

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2004, 03:17:00 ; Search time 2006 Seconds  
(without alignments)  
342.388 Million cell updates/sec

Title: US-10-057-890A-26  
Perfect score: 23  
Sequence: 1 gagtggtggtggtgaccgtgaac 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result			%		Query		DB	ID	Description
			No.	Score	Match	Length			
c	1	20.4	88.7	285	29	CG806452			CG806452 1118070B1
	2	20.4	88.7	358	28	BZ929924			BZ929924 CH240_35I
c	3	20.4	88.7	840	28	BZ826053			BZ826053 PUFY08TD
c	4	19.4	84.3	200	13	BX679676			BX679676 BX679676
c	5	19.4	84.3	299	13	BX680786			BX680786 BX680786
	6	19.4	84.3	325	13	BX681790			BX681790 BX681790
c	7	19.4	84.3	361	13	BX680294			BX680294 BX680294
c	8	19.4	84.3	387	13	BX681122			BX681122 BX681122
c	9	19.4	84.3	397	13	BX681074			BX681074 BX681074
c	10	19.4	84.3	489	13	BX680666			BX680666 BX680666
c	11	19.4	84.3	502	13	BX680451			BX680451 BX680451
c	12	19.4	84.3	508	13	BX680620			BX680620 BX680620
c	13	19.4	84.3	510	13	BX680216			BX680216 BX680216
c	14	19.4	84.3	514	13	BX679987			BX679987 BX679987
c	15	19.4	84.3	515	13	BX681034			BX681034 BX681034
c	16	19.4	84.3	521	13	BX681031			BX681031 BX681031
c	17	19.4	84.3	552	9	AL750676			AL750676 AL750676
c	18	19.4	84.3	563	9	AL751097			AL751097 AL751097
	19	19.4	84.3	921	29	CNS0294E			AL186791 Tetraodon
c	20	19	82.6	314	14	CB682175			CB682175 OSJNEf09D
c	21	19	82.6	607	14	CA754057			CA754057 BR0400080
c	22	19	82.6	701	28	BZ264918			BZ264918 CH230-458
c	23	19	82.6	728	14	CA198100			CA198100 SCEPAM110
	24	19	82.6	762	10	BF617538			BF617538 HVSMEc001
	25	19	82.6	1201	9	AL571982			AL571982 AL571982
	26	18.8	81.7	441	9	AI851068			AI851068 UI-M-BH0-
	27	18.8	81.7	613	29	CE285984			CE285984 tigr-gss-
	28	18.8	81.7	1057	10	BF785786			BF785786 602112408
	29	18.8	81.7	1249	29	CG749526			CG749526 P043-4-B0
	30	18.4	80.0	201	12	BI033996			BI033996 PM2-NN016
	31	18.4	80.0	326	14	CA773745			CA773745 im56e07.y
c	32	18.4	80.0	364	28	BZ330607			BZ330607 hv93h05.b
	33	18.4	80.0	370	14	R34536			R34536 yg58e03.rl
	34	18.4	80.0	414	28	AZ151688			AZ151688 SP_0036_B
c	35	18.4	80.0	457	14	CB877269			CB877269 HP04D04T
c	36	18.4	80.0	467	12	BI367987			BI367987 RE53880.5
	37	18.4	80.0	490	28	AQ953190			AQ953190 Sheared D
c	38	18.4	80.0	512	14	CB878727			CB878727 HP09J14T
	39	18.4	80.0	514	10	BE592946			BE592946 WS1_92_D0
c	40	18.4	80.0	515	29	TA70D06Q			AL458152 T. brucei
	41	18.4	80.0	532	14	CD418822			CD418822 ri77f11.y
	42	18.4	80.0	581	28	BZ451563			BZ451563 BONED12TF
	43	18.4	80.0	616	12	BI753028			BI753028 603025830
c	44	18.4	80.0	623	28	AQ650579			AQ650579 Sheared D
c	45	18.4	80.0	653	28	BH692630			BH692630 BOMNX63TF

## ALIGNMENTS

RESULT 1  
 CG806452/c  
 LOCUS CG806452 285 bp DNA linear GSS 10-NOV-2003  
 DEFINITION 1118070B11.x1 1118 - RescueMu Grid S Zea mays genomic, genomic survey sequence.  
 ACCESSION CG806452  
 VERSION CG806452.1 GI:38243545  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 285)  
 AUTHORS Walbot,V.  
 TITLE Maize genomic sequences found using engineered RescueMu transposon  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Possible ligation site so sequence was trimmed. Post-ligation sequence submitted separately.  
 Plate: 1118070 row: 3  
 Class: transposon-tagged.  
 FEATURES Location/Qualifiers  
 source 1. .285  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="mixed background W23/A188/B73"  
 /db\_xref="taxon:4577"  
 /tissue\_type="leaf"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="1118 - RescueMu Grid S"  
 /note="Organ: leaf; Vector: RescueMu (engineered from pBlueScript backbone); Site\_1: BamHI; Site\_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid S was grown at San Diego in 2002. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

ORIGIN

Query Match 88.7%; Score 20.4; DB 29; Length 285;  
Best Local Similarity 95.5%; Pred. No. 2.8e+03;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAA 22  
||||||| |||||||||  
Db 26 GAGTGGTGATGGTGACCGTGAA 5

RESULT 2

BZ929924

LOCUS BZ929924 358 bp DNA linear GSS 13-JUN-2003

DEFINITION CH240\_35I24.TJ CHORI-240 Bos taurus genomic clone CH240\_35I24,  
genomic survey sequence.

ACCESSION BZ929924

VERSION BZ929924.1 GI:31715303

KEYWORDS GSS.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 358)

AUTHORS Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,  
Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,  
Womack,J.E., de Jong,P.J. and Lewin,H.A.

TITLE A Cattle-Human Comparative Map Built with Cattle BAC-ends and Human  
Genome Sequence

JOURNAL Unpublished (2003)

COMMENT Other\_GSSs: CH240\_35I24.TV

Contact: Harris Lewin

Department of Animal Sciences

University of Illinois at Urbana Champaign

1201 W. Gregory Dr., Urbana, IL 61801, USA

Tel: 217 333 5998

Fax: 217 244 5617

Email: h-lewin@uiuc.edu

Clones are derived from the bovine BAC library CHORI-240

(<http://www.chori.org/bacpac/bovine240.htm>). For BAC library  
availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).

Clones may be purchased from BACPAC Resources

([http://www.chori.org/bacpac/ordering\\_information.htm](http://www.chori.org/bacpac/ordering_information.htm)). This work

was undertaken as part of the International Bovine BAC Mapping

Consortium (IBBMC) by University of Illinois at Urbana

Champaign, USA with funds provided by grant No. AG202-34480-11828

from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock

Genome Sequencing Initiative)

Plate: 35 row: I column: 24

Seq primer: SP6

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1. .358

/organism="Bos taurus"

/mol\_type="genomic DNA"

/strain="breed: Hereford"

/db\_xref="taxon:9913"

/clone="CH240\_35I24"

/sex="Male"  
/cell\_type="Blood"  
/clone\_lib="CHORI-240"  
/note="Vector: pTARBAC1.3; Site\_1: MboI; Site\_2: MboI;  
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC  
library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 88.7%; Score 20.4; DB 28; Length 358;  
Best Local Similarity 95.5%; Pred. No. 2.9e+03;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAAC 23  
|||||||  
Db 310 AGTGGTGGTGGTGACTGTGAAC 331

RESULT 3

BZ826053/c

LOCUS BZ826053 840 bp DNA linear GSS 18-MAR-2003

DEFINITION PUFCY08TD ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBTa291A15,  
genomic survey sequence.

ACCESSION BZ826053

VERSION BZ826053.1 GI:29045174

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 840)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
Bennetzen,J.

TITLE Maize Genomics Consortium

JOURNAL Unpublished (2003)

COMMENT Other\_GSSs: PUFCY08TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES Location/Qualifiers

source

1. .840

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBTa291A15"

/clone\_lib="ZM\_0.6\_1.0\_KB"

/note="Vector: pCR4-TOPO; Site\_1: EcoRI; 0.6-1.0 kb high

CoT selected genomic DNA library"

ORIGIN

Query Match 88.7%; Score 20.4; DB 28; Length 840;  
Best Local Similarity 95.5%; Pred. No. 3.3e+03;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAA 22  
||||||| |||||||||  
Db 749 GAGTGGTGATGGTGACCGTGAA 728

RESULT 4

BX679676/c

LOCUS BX679676 200 bp mRNA linear EST 28-OCT-2003

DEFINITION BX679676 RS Pinus pinaster cDNA clone RS31A01, mRNA sequence.

ACCESSION BX679676

VERSION BX679676.1 GI:38013587

KEYWORDS EST.

SOURCE Pinus pinaster

ORGANISM Pinus pinaster

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE 1 (bases 1 to 200)

AUTHORS Frigerio, J. and Plomion, C.

TITLE Identification of water-deficit responsive genes in Maritime pine  
(Pinus pinaster Ait.) using an EST approach

JOURNAL Unpublished (2002)

COMMENT Contact: Frigerio JM

Genetique et Amelioration 69

INRA

route d'Arcachon 33612 Cestas CEDEX France

Email: Frigerio@pierreton.inra.fr

Email: Frigerio@pierreton.inra.fr

Seq primer: T3.

FEATURES

source

Location/Qualifiers

1..200

/organism="Pinus pinaster"

/mol\_type="mRNA"

/db\_xref="taxon:71647"

/clone="RS31A01"

/tissue\_type="root"

/dev\_stage="6 weeks old seedling"

/lab\_host="SOLR"

/clone\_lib="RS"

/note="Vector: Uni-ZAP XR; ecotype: Landes; The library  
was made from the roots of 6 weeks old seedlings grown in  
hydroponic conditions. A three weeks drought stress  
treatment was applied by lowering the osmotic potential of  
the nutrient solution to -0.45 MPa using 170 g/l of  
polyethylene glycol as an osmoticum. A mixture of  
genotypes were used. Oligo-dT primed cDNA was  
directionally cloned into the EcoRI-XhoI lambda-ZAP vector  
arms and mass-excised to form a pBluescript phagemid"

ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 200;  
Best Local Similarity 95.2%; Pred. No. 5.8e+03;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23  
 |||||||||||||||||  
 Db 165 GTGGTGGTGGTGACCGTGACC 145

RESULT 5

BX680786/c

LOCUS BX680786 299 bp mRNA linear EST 28-OCT-2003

DEFINITION BX680786 RS Pinus pinaster cDNA clone RS48D09, mRNA sequence.

ACCESSION BX680786

VERSION BX680786.1 GI:38015244

KEYWORDS EST.

SOURCE Pinus pinaster

ORGANISM Pinus pinaster

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE 1 (bases 1 to 299)

AUTHORS Frigerio, J. and Plomion, C.

TITLE Identification of water-deficit responsive genes in Maritime pine  
 (Pinus pinaster Ait.) using an EST approach

JOURNAL Unpublished (2002)

COMMENT Contact: Frigerio JM  
 Genetique et Amelioration 69

INRA

route d'Arcachon 33612 Cestas CEDEX France

Email: Frigerio@pierreton.inra.fr

Email: Frigerio@pierreton.inra.fr

Seq primer: T3.

FEATURES

source

Location/Qualifiers

1. .299

/organism="Pinus pinaster"

/mol\_type="mRNA"

/db\_xref="taxon:71647"

/clone="RS48D09"

/tissue\_type="root"

/dev\_stage="6 weeks old seedling"

/lab\_host="SOLR"

/clone\_lib="RS"

/note="Vector: Uni-ZAP XR; ecotype: Landes; The library  
 was made from the roots of 6 weeks old seedlings grown in  
 hydroponic conditions. A three weeks drought stress  
 treatment was applied by lowering the osmotic potential of  
 the nutrient solution to -0.45 MPa using 170 g/l of  
 polyethylene glycol as an osmoticum. A mixture of  
 genotypes were used. Oligo-dT primed cDNA was  
 directionally cloned into the EcoRI-XhoI lambda-ZAP vector  
 arms and mass-excised to form a pBluescript phagemid"

ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 299;

Best Local Similarity 95.2%; Pred. No. 6.2e+03;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23

|||||||||||||||||

Db 203 GTGGTGGTGGTGACCGTGACC 183

RESULT 6

BX681790

LOCUS BX681790 325 bp mRNA linear EST 04-NOV-2003

DEFINITION BX681790 RS Pinus pinaster cDNA clone RS64H09, mRNA sequence.

ACCESSION BX681790

VERSION BX681790.1 GI:38158002

KEYWORDS EST.

SOURCE Pinus pinaster

ORGANISM Pinus pinaster

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE 1 (bases 1 to 325)

AUTHORS Frigerio, J. and Plomion, C.

TITLE Identification of water-deficit responsive genes in Maritime pine  
(Pinus pinaster Ait.) using an EST approach

JOURNAL Unpublished (2002)

COMMENT Contact: Frigerio JM

Genetique et Amelioration 69

INRA

route d'Arcachon 33612 Cestas CEDEX France

Email: Frigerio@pierreton.inra.fr

Email: Frigerio@pierreton.inra.fr

Seq primer: T3.

FEATURES Location/Qualifiers

source

1. .325

/organism="Pinus pinaster"

/mol\_type="mRNA"

/db\_xref="taxon:71647"

/clone="RS64H09"

/tissue\_type="root"

/dev\_stage="6 weeks old seedling"

/lab\_host="SOLR"

/clone\_lib="RS"

/note="Vector: Uni-ZAP XR; ecotype: Landes; The library  
was made from the roots of 6 weeks old seedlings grown in  
hydroponic conditions. A three weeks drought stress  
treatment was applied by lowering the osmotic potential of  
the nutrient solution to -0.45 MPa using 170 g/l of  
polyethylene glycol as an osmoticum. A mixture of  
genotypes were used. Oligo-dT primed cDNA was  
directionally cloned into the EcoRI-XhoI lambda-ZAP vector  
arms and mass-excised to form a pBluescript phagemid"

ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 325;

Best Local Similarity 95.2%; Pred. No. 6.3e+03;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23

|||||||

Db 257 GTGGTGGTGGTGACCGTGACC 277

RESULT 7

BX680294/c

LOCUS BX680294 361 bp mRNA linear EST 28-OCT-2003



DEFINITION BX680294 RS Pinus pinaster cDNA clone RS41F04, mRNA sequence.  
 ACCESSION BX680294  
 VERSION BX680294.1 GI:38014421  
 KEYWORDS EST.  
 SOURCE Pinus pinaster  
 ORGANISM Pinus pinaster  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
 REFERENCE 1 (bases 1 to 361)  
 AUTHORS Frigerio, J. and Plomion, C.  
 TITLE Identification of water-deficit responsive genes in Maritime pine  
 (Pinus pinaster Ait.) using an EST approach  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Frigerio JM  
 Genetique et Amelioration 69  
 INRA  
 route d'Arcachon 33612 Cestas CEDEX France  
 Email: Frigerio@pierroton.inra.fr  
 Email: Frigerio@pierroton.inra.fr  
 Seq primer: T3.  
 FEATURES Location/Qualifiers  
 source 1..361  
 /organism="Pinus pinaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:71647"  
 /clone="RS41F04"  
 /tissue\_type="root"  
 /dev\_stage="6 weeks old seedling"  
 /lab\_host="SOLR"  
 /clone\_lib="RS"  
 /note="Vector: Uni-ZAP XR; ecotype: Landes; The library  
 was made from the roots of 6 weeks old seedlings grown in  
 hydroponic conditions. A three weeks drought stress  
 treatment was applied by lowering the osmotic potential of  
 the nutrient solution to -0.45 MPa using 170 g/l of  
 polyethylene glycol as an osmoticum. A mixture of  
 genotypes were used. Oligo-dT primed cDNA was  
 directionally cloned into the EcoRI-XhoI lambda-ZAP vector  
 arms and mass-excised to form a pBluescript phagemid"

# ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 361;  
 Best Local Similarity 95.2%; Pred. No. 6.4e+03;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23  
 |||||  
 Db 151 GTGGTGGTGGTGACCGTGACC 131

RESULT 8  
 BX681122/c  
 LOCUS BX681122 387 bp mRNA linear EST 28-OCT-2003  
 DEFINITION BX681122 RS Pinus pinaster cDNA clone RS53H04, mRNA sequence.  
 ACCESSION BX681122  
 VERSION BX681122.1 GI:38015580  
 KEYWORDS EST.

SOURCE Pinus pinaster  
 ORGANISM Pinus pinaster  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
 REFERENCE 1 (bases 1 to 387)  
 AUTHORS Frigerio, J. and Plomion, C.  
 TITLE Identification of water-deficit responsive genes in Maritime pine  
 (Pinus pinaster Ait.) using an EST approach  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Frigerio JM  
 Genetique et Amelioration 69  
 INRA  
 route d'Arcachon 33612 Cestas CEDEX France  
 Email: Frigerio@pierroton.inra.fr  
 Email: Frigerio@pierroton.inra.fr  
 Seq primer: T3.

FEATURES Location/Qualifiers  
 source 1. .387  
 /organism="Pinus pinaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:71647"  
 /clone="RS53H04"  
 /tissue\_type="root"  
 /dev\_stage="6 weeks old seedling"  
 /lab\_host="SOLR"  
 /clone\_lib="RS"  
 /note="Vector: Uni-ZAP XR; ecotype: Landes; The library  
 was made from the roots of 6 weeks old seedlings grown in  
 hydroponic conditions. A three weeks drought stress  
 treatment was applied by lowering the osmotic potential of  
 the nutrient solution to -0.45 MPa using 170 g/l of  
 polyethylene glycol as an osmoticum. A mixture of  
 genotypes were used. Oligo-dT primed cDNA was  
 directionally cloned into the EcoRI-XhoI lambda-ZAP vector  
 arms and mass-excised to form a pBluescript phagemid"

#### ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 387;  
 Best Local Similarity 95.2%; Pred. No. 6.5e+03;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23  
 |||||  
 Db 150 GTGGTGGTGGTGACCGTGACC 130

#### RESULT 9

BX681074/c

LOCUS BX681074 397 bp mRNA linear EST 04-NOV-2003  
 DEFINITION BX681074 RS Pinus pinaster cDNA clone RS52H04, mRNA sequence.  
 ACCESSION BX681074  
 VERSION BX681074.1 GI:38015532  
 KEYWORDS EST.  
 SOURCE Pinus pinaster  
 ORGANISM Pinus pinaster  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE 1 (bases 1 to 397)  
AUTHORS Frigerio, J. and Plomion, C.  
TITLE Identification of water-deficit responsive genes in Maritime pine  
(Pinus pinaster Ait.) using an EST approach  
JOURNAL Unpublished (2002)  
COMMENT Contact: Frigerio JM  
Genetique et Amelioration 69  
INRA  
route d'Arcachon 33612 Cestas CEDEX France  
Email: Frigerio@pierreton.inra.fr  
Email: Frigerio@pierreton.inra.fr  
Seq primer: T3.

FEATURES Location/Qualifiers  
source 1. .397  
/organism="Pinus pinaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:71647"  
/clone="RS52H04"  
/tissue\_type="root"  
/dev\_stage="6 weeks old seedling"  
/lab\_host="SOLR"  
/clone\_lib="RS"  
/note="Vector: Uni-ZAP XR; ecotype: Landes; The library  
was made from the roots of 6 weeks old seedlings grown in  
hydroponic conditions. A three weeks drought stress  
treatment was applied by lowering the osmotic potential of  
the nutrient solution to -0.45 MPa using 170 g/l of  
polyethylene glycol as an osmoticum. A mixture of  
genotypes were used. Oligo-dT primed cDNA was  
directionally cloned into the EcoRI-XhoI lambda-ZAP vector  
arms and mass-excised to form a pBluescript phagemid"

# ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 397;  
Best Local Similarity 95.2%; Pred. No. 6.5e+03;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23  
|||||||  
Db 159 GTGGTGGTGGTGACCGTGACC 139

RESULT 10  
BX680666/c  
LOCUS BX680666 489 bp mRNA linear EST 28-OCT-2003  
DEFINITION BX680666 RS Pinus pinaster cDNA clone RS47A03, mRNA sequence.  
ACCESSION BX680666  
VERSION BX680666.1 GI:38015124  
KEYWORDS EST.  
SOURCE Pinus pinaster  
ORGANISM Pinus pinaster  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
REFERENCE 1 (bases 1 to 489)  
AUTHORS Frigerio, J. and Plomion, C.  
TITLE Identification of water-deficit responsive genes in Maritime pine  
(Pinus pinaster Ait.) using an EST approach

JOURNAL Unpublished (2002)  
COMMENT Contact: Frigerio JM  
Genetique et Amelioration 69  
INRA  
route d'Arcachon 33612 Cestas CEDEX France  
Email: Frigerio@pierreton.inra.fr  
Email: Frigerio@pierreton.inra.fr  
Seq primer: T3.

FEATURES Location/Qualifiers  
source 1. 489  
/organism="Pinus pinaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:71647"  
/clone="RS47A03"  
/tissue\_type="root"  
/dev\_stage="6 weeks old seedling"  
/lab\_host="SOLR"  
/clone\_lib="RS"  
/note="Vector: Uni-ZAP XR; ecotype: Landes; The library was made from the roots of 6 weeks old seedlings grown in hydroponic conditions. A three weeks drought stress treatment was applied by lowering the osmotic potential of the nutrient solution to -0.45 MPa using 170 g/l of polyethylene glycol as an osmoticum. A mixture of genotypes were used. Oligo-dT primed cDNA was directionally cloned into the EcoRI-XhoI lambda-ZAP vector arms and mass-excised to form a pBluescript phagemid"

ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 489;  
Best Local Similarity 95.2%; Pred. No. 6.7e+03;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GTGGTGGTGGTGACCGTGAAC 23  
|||||||  
Db 187 GTGGTGGTGGTGACCGTGACC 167

RESULT 11  
BX680451/c

LOCUS BX680451 502 bp mRNA linear EST 28-OCT-2003  
DEFINITION BX680451 RS Pinus pinaster cDNA clone RS43E08, mRNA sequence.  
ACCESSION BX680451  
VERSION BX680451.1 GI:38014907  
KEYWORDS EST.  
SOURCE Pinus pinaster  
ORGANISM Pinus pinaster  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
REFERENCE 1 (bases 1 to 502)  
AUTHORS Frigerio, J. and Plomion, C.  
TITLE Identification of water-deficit responsive genes in Maritime pine (Pinus pinaster Ait.) using an EST approach  
JOURNAL Unpublished (2002)  
COMMENT Contact: Frigerio JM  
Genetique et Amelioration 69  
INRA

route d'Arcachon 33612 Cestas CEDEX France  
Email: Frigerio@pierroton.inra.fr  
Email: Frigerio@pierroton.inra.fr  
Seq primer: T3.

FEATURES  
    source                   Location/Qualifiers  
                            1. .502  
                            /organism="Pinus pinaster"  
                            /mol\_type="mRNA"  
                            /db\_xref="taxon:71647"  
                            /clone="RS43E08"  
                            /tissue\_type="root"  
                            /dev\_stage="6 weeks old seedling"  
                            /lab\_host="SOLR"  
                            /clone\_lib="RS"  
                            /note="Vector: Uni-ZAP XR; ecotype: Landes; The library  
                            was made from the roots of 6 weeks old seedlings grown in  
                            hydroponic conditions. A three weeks drought stress  
                            treatment was applied by lowering the osmotic potential of  
                            the nutrient solution to -0.45 MPa using 170 g/l of  
                            polyethylene glycol as an osmoticum. A mixture of  
                            genotypes were used. Oligo-dT primed cDNA was  
                            directionally cloned into the EcoRI-XhoI lambda-ZAP vector  
                            arms and mass-excised to form a pBluescript phagemid"

ORIGIN

Query Match                   84.3%;   Score 19.4;   DB 13;   Length 502;  
Best Local Similarity   95.2%;   Pred. No. 6.8e+03;  
Matches   20;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy           3 GTGGTGGTGGTGACCGTGAAC 23  
             | | | | | | | | | | | | | | | |  
Db           145 GTGGTGGTGGTGACCGTGACC 125

RESULT 12  
BX680620/c  
LOCUS           BX680620                   508 bp    mRNA    linear   EST 28-OCT-2003  
DEFINITION   BX680620 RS Pinus pinaster cDNA clone RS46A03, mRNA sequence.  
ACCESSION    BX680620  
VERSION      BX680620.1   GI:38015078  
KEYWORDS     EST.  
SOURCE       Pinus pinaster  
    ORGANISM   Pinus pinaster  
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
              Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
REFERENCE    1 (bases 1 to 508)  
    AUTHORS   Frigerio, J. and Plomion, C.  
    TITLE     Identification of water-deficit responsive genes in Maritime pine  
              (Pinus pinaster Ait.) using an EST approach  
    JOURNAL   Unpublished (2002)  
COMMENT      Contact: Frigerio JM  
              Genetique et Amelioration 69  
              INRA  
              route d'Arcachon 33612 Cestas CEDEX France  
              Email: Frigerio@pierroton.inra.fr  
              Email: Frigerio@pierroton.inra.fr  
              Seq primer: T3.

FEATURES  
 source                      Location/Qualifiers  
                              1. .508  
                              /organism="Pinus pinaster"  
                              /mol\_type="mRNA"  
                              /db\_xref="taxon:71647"  
                              /clone="RS46A03"  
                              /tissue\_type="root"  
                              /dev\_stage="6 weeks old seedling"  
                              /lab\_host="SOLR"  
                              /clone\_lib="RS"  
                              /note="Vector: Uni-ZAP XR; ecotype: Landes; The library  
                              was made from the roots of 6 weeks old seedlings grown in  
                              hydroponic conditions. A three weeks drought stress  
                              treatment was applied by lowering the osmotic potential of  
                              the nutrient solution to -0.45 MPa using 170 g/l of  
                              polyethylene glycol as an osmoticum. A mixture of  
                              genotypes were used. Oligo-dT primed cDNA was  
                              directionally cloned into the EcoRI-XhoI lambda-ZAP vector  
                              arms and mass-excised to form a pBluescript phagemid"

ORIGIN

Query Match                      84.3%;   Score 19.4;   DB 13;   Length 508;  
 Best Local Similarity    95.2%;   Pred. No. 6.8e+03;  
 Matches    20;   Conservative    0;   Mismatches    1;   Indels       0;   Gaps       0;

Qy                      3 GTGGTGGTGGTGACCGTGAAC 23  
                              |||||  
 Db                      178 GTGGTGGTGGTGACCGTGACC 158

RESULT 13

BX680216/c

LOCUS                      BX680216                      510 bp       mRNA       linear       EST 28-OCT-2003

DEFINITION    BX680216 RS Pinus pinaster cDNA clone RS40F04, mRNA sequence.

ACCESSION    BX680216

VERSION       BX680216.1    GI:38014264

KEYWORDS       EST.

SOURCE       Pinus pinaster

ORGANISM    Pinus pinaster

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE    1 (bases 1 to 510)

AUTHORS       Frigerio, J. and Plomion, C.

TITLE       Identification of water-deficit responsive genes in Maritime pine  
 (Pinus pinaster Ait.) using an EST approach

JOURNAL       Unpublished (2002)

COMMENT       Contact: Frigerio JM

Genetique et Amelioration 69

INRA

route d'Arcachon 33612 Cestas CEDEX France

Email: Frigerio@pierreton.inra.fr

Email: Frigerio@pierreton.inra.fr

Seq primer: T3.

FEATURES                      Location/Qualifiers

source                      1. .510

/organism="Pinus pinaster"

/mol\_type="mRNA"

/db\_xref="taxon:71647"  
 /clone="RS40F04"  
 /tissue\_type="root"  
 /dev\_stage="6 weeks old seedling"  
 /lab\_host="SOLR"  
 /clone\_lib="RS"  
 /note="Vector: Uni-ZAP XR; ecotype: Landes; The library was made from the roots of 6 weeks old seedlings grown in hydroponic conditions. A three weeks drought stress treatment was applied by lowering the osmotic potential of the nutrient solution to -0.45 MPa using 170 g/l of polyethylene glycol as an osmoticum. A mixture of genotypes were used. Oligo-dT primed cDNA was directionally cloned into the EcoRI-XhoI lambda-ZAP vector arms and mass-excised to form a pBluescript phagemid"

# ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 510;  
 Best Local Similarity 95.2%; Pred. No. 6.8e+03;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23  
 |||||  
 Db 168 GTGGTGGTGGTGACCGTGACC 148

## RESULT 14 BX679987/c

LOCUS BX679987 514 bp mRNA linear EST 28-OCT-2003  
 DEFINITION BX679987 RS Pinus pinaster cDNA clone RS37G01, mRNA sequence.  
 ACCESSION BX679987  
 VERSION BX679987.1 GI:38013898  
 KEYWORDS EST.  
 SOURCE Pinus pinaster  
 ORGANISM Pinus pinaster  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
 REFERENCE 1 (bases 1 to 514)  
 AUTHORS Frigerio, J. and Plomion, C.  
 TITLE Identification of water-deficit responsive genes in Maritime pine (Pinus pinaster Ait.) using an EST approach  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Frigerio JM  
 Genetique et Amelioration 69  
 INRA  
 route d'Arcachon 33612 Cestas CEDEX France  
 Email: Frigerio@pierreton.inra.fr  
 Email: Frigerio@pierreton.inra.fr  
 Seq primer: T3.

FEATURES Location/Qualifiers  
 source 1..514  
 /organism="Pinus pinaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:71647"  
 /clone="RS37G01"  
 /tissue\_type="root"  
 /dev\_stage="6 weeks old seedling"

```

/lab_host="SOLR"
/clone_lib="RS"
/note="Vector: Uni-ZAP XR; ecotype: Landes; The library
was made from the roots of 6 weeks old seedlings grown in
hydroponic conditions. A three weeks drought stress
treatment was applied by lowering the osmotic potential of
the nutrient solution to -0.45 MPa using 170 g/l of
polyethylene glycol as an osmoticum. A mixture of
genotypes were used. Oligo-dT primed cDNA was
directionally cloned into the EcoRI-XhoI lambda-ZAP vector
arms and mass-excised to form a pBluescript phagemid"

```

# ORIGIN

```

Query Match          84.3%;  Score 19.4;  DB 13;  Length 514;
Best Local Similarity 95.2%;  Pred. No. 6.8e+03;
Matches   20;  Conservative   0;  Mismatches   1;  Indels       0;  Gaps       0;

```

```

QY          3 GTGGTGGTGGTGACCGTGAAC 23
             |||||
Db          209 GTGGTGGTGGTGACCGTGACC 189

```

## RESULT 15

BX681034/c

LOCUS BX681034 515 bp mRNA linear EST 28-OCT-2003

DEFINITION BX681034 RS Pinus pinaster cDNA clone RS52C06, mRNA sequence.

ACCESSION BX681034

VERSION BX681034.1 GI:38015492

KEYWORDS EST.

SOURCE Pinus pinaster

ORGANISM Pinus pinaster

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE 1 (bases 1 to 515)

AUTHORS Frigerio, J. and Plomion, C.

TITLE Identification of water-deficit responsive genes in Maritime pine  
(Pinus pinaster Ait.) using an EST approach

JOURNAL Unpublished (2002)

COMMENT Contact: Frigerio JM

Genetique et Amelioration 69

INRA

route d'Arcachon 33612 Cestas CEDEX France

Email: Frigerio@pierreton.inra.fr

Email: Frigerio@pierreton.inra.fr

Seq primer: T3.

## FEATURES

source

Location/Qualifiers

1..515

/organism="Pinus pinaster"

/mol\_type="mRNA"

/db\_xref="taxon:71647"

/clone="RS52C06"

/tissue\_type="root"

/dev\_stage="6 weeks old seedling"

/lab\_host="SOLR"

/clone\_lib="RS"

/note="Vector: Uni-ZAP XR; ecotype: Landes; The library

was made from the roots of 6 weeks old seedlings grown in



hydroponic conditions. A three weeks drought stress treatment was applied by lowering the osmotic potential of the nutrient solution to -0.45 MPa using 170 g/l of polyethylene glycol as an osmoticum. A mixture of genotypes were used. Oligo-dT primed cDNA was directionally cloned into the EcoRI-XhoI lambda-ZAP vector arms and mass-excised to form a pBluescript phagemid"

ORIGIN

Query Match 84.3%; Score 19.4; DB 13; Length 515;  
Best Local Similarity 95.2%; Pred. No. 6.8e+03;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23  
| | | | | | | | | | | | | | | | | |  
Db 219 GTGGTGGTGGTGACCGTGACC 199

Search completed: March 11, 2004, 08:07:25  
Job time : 2011 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 10, 2004, 22:19:04 ; Search time 1677 Seconds  
(without alignments)  
594.448 Million cell updates/sec

Title: US-10-057-890A-26  
Perfect score: 23  
Sequence: 1 gagtggtggtggtgaccgtgaac 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*

28: em\_un:\*  
 29: em\_vi:\*  
 30: em\_htg\_hum:\*  
 31: em\_htg\_inv:\*  
 32: em\_htg\_other:\*  
 33: em\_htg\_mus:\*  
 34: em\_htg\_pln:\*  
 35: em\_htg\_rod:\*  
 36: em\_htg\_mam:\*  
 37: em\_htg\_vrt:\*  
 38: em\_sy:\*  
 39: em\_htgo\_hum:\*  
 40: em\_htgo\_mus:\*  
 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result			Query				Description
	No.	Score	Match	Length	DB	ID	
c	1	20.4	88.7	228931	2	AC131355	AC131355 Rattus no
	2	20.4	88.7	257619	2	AC114512	AC114512 Rattus no
	3	19.8	86.1	140416	9	AC013630	AC013630 Homo sapi
c	4	19.8	86.1	175569	2	AP001151	AP001151 Homo sapi
c	5	19.8	86.1	197834	2	AP002434	AP002434 Homo sapi
c	6	19.8	86.1	206736	2	AP001200	AP001200 Homo sapi
	7	19.8	86.1	214984	9	AC015563	AC015563 Homo sapi
c	8	19.4	84.3	2000	6	AX655151	AX655151 Sequence
c	9	19.4	84.3	148892	8	AP003235	AP003235 Oryza sat
c	10	19.4	84.3	152736	8	AP003566	AP003566 Oryza sat
c	11	19	82.6	460	8	AY117122	AY117122 Rhizopogo
c	12	19	82.6	2638	10	AB096080	AB096080 Rattus no
c	13	19	82.6	2657	10	AF109963	AF109963 Rattus no
	14	19	82.6	99995	9	AC010480	AC010480 Homo sapi
c	15	19	82.6	142325	2	AP004191	AP004191 Oryza sat
	16	19	82.6	155939	8	AP003853	AP003853 Oryza sat
c	17	19	82.6	173902	2	AC024667	AC024667 Homo sapi
c	18	19	82.6	175517	9	AC096586	AC096586 Homo sapi
c	19	19	82.6	201404	9	AC020728	AC020728 Homo sapi
c	20	19	82.6	282899	2	AC095879	AC095879 Rattus no
c	21	18.8	81.7	1083	8	AF290565	AF290565 Brassica
c	22	18.8	81.7	1561	8	AF290568	AF290568 Brassica
	23	18.8	81.7	216989	2	AC137155	AC137155 Mus muscu
c	24	18.8	81.7	229511	2	AC119666	AC119666 Rattus no
	25	18.8	81.7	230278	14	MCU68299	U68299 Mouse cytom
	26	18.8	81.7	234117	2	AC130985	AC130985 Rattus no
	27	18.4	80.0	2295	6	BD130230	BD130230 Human sig
	28	18.4	80.0	2332	9	BC009381	BC009381 Homo sapi
	29	18.4	80.0	35023	2	AC141234	AC141234 Homo sapi
	30	18.4	80.0	38005	2	AC140705	AC140705 Homo sapi
c	31	18.4	80.0	39921	2	AC141237	AC141237 Homo sapi
	32	18.4	80.0	63952	9	AL441883	AL441883 Human DNA
	33	18.4	80.0	89642	2	AC005136	AC005136 Homo sapi

	34	18.4	80.0	109423	9	AC027275	AC027275 Homo sapi
c	35	18.4	80.0	110000	2	AC130075_1	Continuation (2 of
c	36	18.4	80.0	114793	9	AF217796	AF217796 Homo sapi
c	37	18.4	80.0	119696	2	AP005563	AP005563 Oryza sat
c	38	18.4	80.0	122753	8	AC124971	AC124971 Medicago
c	39	18.4	80.0	123166	2	AC140913	AC140913 Medicago
c	40	18.4	80.0	124048	9	HUAC002045	AC002045 Human Chr
c	41	18.4	80.0	124437	2	AC141598	AC141598 Homo sapi
	42	18.4	80.0	127485	9	HUAC002039	AC002039 Homo sapi
c	43	18.4	80.0	127925	9	AC135593	AC135593 Homo sapi
	44	18.4	80.0	132781	2	AC141265	AC141265 Homo sapi
	45	18.4	80.0	133726	2	AC141614	AC141614 Homo sapi

# ALIGNMENTS

## RESULT 1

AC131355/c

LOCUS AC131355 228931 bp DNA linear HTG 11-OCT-2002

DEFINITION Rattus norvegicus clone CH230-289D4, WORKING DRAFT SEQUENCE, 2 unordered pieces.

ACCESSION AC131355

VERSION AC131355.2 GI:23603734

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 228931)

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hagues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Louseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,  
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 228931)  
AUTHORS Rat Genome Sequencing Consortium.  
TITLE Direct Submission  
JOURNAL Submitted (21-AUG-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
REFERENCE 3 (bases 1 to 228931)  
AUTHORS Rat Genome Sequencing Consortium.  
TITLE Direct Submission  
JOURNAL Submitted (11-OCT-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
COMMENT On Oct 9, 2002 this sequence version replaced gi:22380619.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GPIN

Center clone name: CH230-289D4

----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 210030 bases at least Q40  
Consensus quality: 213596 bases at least Q30  
Consensus quality: 216121 bases at least Q20  
Estimated insert size: 220403; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 227459: contig of 227459 bp in length  
\* 227460 227559: gap of unknown length  
\* 227560 228931: contig of 1372 bp in length.

FEATURES Location/Qualifiers  
source 1. .228931  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-289D4"  
misc\_feature 1. .1267  
/note="wgs\_end\_extension  
clone\_end:T7"  
misc\_feature 4893. .5714  
/note="clone\_boundary  
clone\_end:T7  
site:MboI  
end\_sequence:RXAID14TJ"  
misc\_feature 220510. .221213  
/note="clone\_boundary  
clone\_end:Sp6  
site:MboI  
end\_sequence:RXAID14TV"  
misc\_feature 225075. .227459  
/note="wgs\_end\_extension  
clone\_end:Sp6"

ORIGIN

Query Match 88.7%; Score 20.4; DB 2; Length 228931;  
Best Local Similarity 95.5%; Pred. No. 2e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAAC 23  
|||||  
Db 98321 AGTGGTGGTGGTGACCGTGCAC 98300

RESULT 2  
AC114512

LOCUS AC114512 257619 bp DNA linear HTG 23-NOV-2002  
 DEFINITION Rattus norvegicus clone CH230-5N9, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 5 unordered pieces.  
 ACCESSION AC114512  
 VERSION AC114512.4 GI:25188484  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 257619)  
 AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hagues,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,

Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 257619)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
REFERENCE 3 (bases 1 to 257619)  
AUTHORS Rat Genome Sequencing Consortium.  
TITLE Direct Submission  
JOURNAL Submitted (23-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
COMMENT On Nov 23, 2002 this sequence version replaced gi:23195170.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GRZR  
Center clone name: CH230-5N9  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 214424 bases at least Q40  
Consensus quality: 219315 bases at least Q30  
Consensus quality: 222843 bases at least Q20  
Estimated insert size: 221074; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.



\* 1 251733: contig of 251733 bp in length  
 \* 251734 251833: gap of unknown length  
 \* 251834 252850: contig of 1017 bp in length  
 \* 252851 252950: gap of unknown length  
 \* 252951 254605: contig of 1655 bp in length  
 \* 254606 254705: gap of unknown length  
 \* 254706 255949: contig of 1244 bp in length  
 \* 255950 256049: gap of unknown length  
 \* 256050 257619: contig of 1570 bp in length.

FEATURES Location/Qualifiers  
 source 1. .257619  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-5N9"  
 misc\_feature 1. .1043  
 /note="wgs\_contig"  
 misc\_feature 144868. .146770  
 /note="wgs\_contig"  
 misc\_feature 193533. .194883  
 /note="wgs\_contig"

# ORIGIN

Query Match 88.7%; Score 20.4; DB 2; Length 257619;  
 Best Local Similarity 95.5%; Pred. No. 2e+02;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAAC 23  
 |||||  
 Db 24969 AGTGGTGGTGGTGACCGTGCAC 24990

# RESULT 3

AC013630

LOCUS AC013630 140416 bp DNA linear PRI 11-DEC-2001

DEFINITION Homo sapiens, clone RP11-12F2, complete sequence.

ACCESSION AC013630

VERSION AC013630.12 GI:17488687

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 140416)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens, clone RP11-12F2

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 140416)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,

	Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
TITLE	Direct Submission
JOURNAL	Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	3 (bases 1 to 140416)
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavkiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	4 (bases 1 to 140416)
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavkiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
 JOURNAL Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Dec 11, 2001 this sequence version replaced gi:14589624.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L3262  
 Center clone name: 12\_F\_2  
 -----  
 FEATURES Location/Qualifiers  
 source 1. .140416  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="RP11-12F2"  
 /clone\_lib="RPCI-11 Human Male BAC"  
 repeat\_region 1446. .1612  
 /rpt\_family="MIR"  
 repeat\_region 3170. .3466  
 /rpt\_family="AluY"  
 repeat\_region 3571. .3616  
 /rpt\_family="(TA)n"  
 repeat\_region 3994. .4136  
 /rpt\_family="MIR"  
 repeat\_region 4352. .4481  
 /rpt\_family="FLAM\_C"  
 repeat\_region complement(6133. .6448)  
 /rpt\_family="LTR16A"  
 repeat\_region 7399. .7428  
 /rpt\_family="(CA)n"  
 repeat\_region complement(8426. .9181)  
 /rpt\_family="Charliela"  
 repeat\_region 9182. .9454  
 /rpt\_family="AluJo"  
 repeat\_region complement(9455. .9654)  
 /rpt\_family="Charliela"  
 repeat\_region complement(9655. .10365)  
 /rpt\_family="L1MD1"  
 repeat\_region complement(10366. .10807)  
 /rpt\_family="L1MC1"  
 repeat\_region complement(10808. .10985)  
 /rpt\_family="L1MD1"  
 unsure 10996  
 /note="probably T"  
 repeat\_region complement(11005. .11178)  
 /rpt\_family="AluSg/x"  
 repeat\_region complement(11185. .11741)  
 /rpt\_family="L1MD1"  
 repeat\_region complement(11747. .12188)  
 /rpt\_family="L1MD1"

repeat_region	complement(12189. .12495) /rpt_family="AluSx"
repeat_region	complement(12496. .13195) /rpt_family="L1MD1"
repeat_region	complement(13196. .13501) /rpt_family="AluSq"
repeat_region	complement(13502. .13835) /rpt_family="L1MD1"
repeat_region	complement(13853. .14044) /rpt_family="L1MC1"
repeat_region	complement(14054. .14208) /rpt_family="Charliela"
repeat_region	14321. .14500 /rpt_family="MIR"
repeat_region	14575. .14640 /rpt_family="MER94"
repeat_region	14641. .14943 /rpt_family="AluSx"
repeat_region	14944. .14998 /rpt_family="MER94"
repeat_region	15060. .15285 /rpt_family="MIR"
repeat_region	15463. .15662 /rpt_family="THE1C"
repeat_region	complement(15673. .21531) /rpt_family="L1PA3"
repeat_region	21532. .21730 /rpt_family="THE1C"
repeat_region	complement(23669. .23782) /rpt_family="MIR"
repeat_region	25137. .25434 /rpt_family="AluJo"
repeat_region	25459. .25602 /rpt_family="FLAM_C"
repeat_region	complement(26027. .26064) /rpt_family="MIR"
repeat_region	complement(26243. .26691) /rpt_family="MLT1K"
repeat_region	26768. .26788 /rpt_family="AT_rich"
repeat_region	26789. .27056 /rpt_family="AluSq"
repeat_region	27057. .27077 /rpt_family="AT_rich"
repeat_region	complement(27504. .27668) /rpt_family="FAM"
repeat_region	27818. .28159 /rpt_family="L1M4"
repeat_region	29097. .29547 /rpt_family="MER4C"
repeat_region	30008. .30577 /rpt_family="MER34B"
repeat_region	30578. .31021 /rpt_family="MLT1C"
repeat_region	31043. .31132 /rpt_family="L2"
repeat_region	31908. .31935

```

repeat_region      /rpt_family="(TTCA)n"
                   complement(32953. .33258)
                   /rpt_family="AluJb"
repeat_region      complement(33478. .33778)
                   /rpt_family="AluSx"
repeat_region      33861. .33888

```

```

Query Match          86.1%;  Score 19.8;  DB 9;  Length 140416;
Best Local Similarity 91.3%;  Pred. No. 3.8e+02;
Matches 21;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

```

```

QY      1 GAGTGGTGGTGGTGACCGTGAAC 23
        |||||
Db      82434 GAGTGGTGGTGGTGACAGTGAGC 82456

```

#### RESULT 4

AP001151/c

LOCUS AP001151 175569 bp DNA linear HTG 30-MAY-2000

DEFINITION Homo sapiens chromosome 18 clone RP11-810016 map 18q12, WORKING  
DRAFT SEQUENCE, 37 unordered pieces.

ACCESSION AP001151

VERSION AP001151.2 GI:8118700

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 175569)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens 175,569 genomic DNA of 18q12

JOURNAL Published Only in DataBase (2000)

REFERENCE 2 (bases 1 to 175569)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (08-FEB-2000) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
Japan (E-mail:hattori@gsc.riken.go.jp,  
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,  
Fax:81-42-778-9924)

COMMENT On May 31, 2000 this sequence version replaced gi:6997829.

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-810016

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 153725 bases at least Q40

Consensus quality: 164241 bases at least Q30  
 Consensus quality: 169066 bases at least Q20  
 Insert size: 171969; sum-of-contigs  
 Quality coverage: 4.80x in Q20 bases; sum-of-contigs

-----  
 NOTE: This is a 'working draft' sequence. It currently consists of 37 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	14201	contig of	14201 bp in length
14302	28801	contig of	14500 bp in length
28902	40910	contig of	12009 bp in length
41011	54579	contig of	13569 bp in length
54680	62489	contig of	7810 bp in length
62590	72710	contig of	10121 bp in length
72811	81356	contig of	8546 bp in length
81457	87558	contig of	6102 bp in length
87659	95328	contig of	7670 bp in length
95429	102558	contig of	7130 bp in length
102659	107119	contig of	4461 bp in length
107220	112453	contig of	5234 bp in length
112554	117698	contig of	5145 bp in length
117799	122290	contig of	4492 bp in length
122391	126549	contig of	4159 bp in length
126650	128718	contig of	2069 bp in length
128819	132710	contig of	3892 bp in length
132811	136147	contig of	3337 bp in length
136248	139746	contig of	3499 bp in length
139847	142112	contig of	2266 bp in length
142213	144472	contig of	2260 bp in length
144573	147200	contig of	2628 bp in length
147301	150373	contig of	3073 bp in length
150474	152540	contig of	2067 bp in length
152641	155258	contig of	2618 bp in length
155359	158220	contig of	2862 bp in length
158321	160464	contig of	2144 bp in length
160565	163294	contig of	2730 bp in length
163395	164687	contig of	1293 bp in length
164788	166129	contig of	1342 bp in length
166230	167303	contig of	1074 bp in length
167404	168688	contig of	1285 bp in length
168789	170602	contig of	1814 bp in length
170703	171966	contig of	1264 bp in length
172067	173286	contig of	1220 bp in length
173387	174411	contig of	1025 bp in length
174512	175569	contig of	1058 bp in length

Sequence updated (26-May-2000).

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 37 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will

\* be preserved.

* 1	14201: contig of 14201 bp in length
* 14202	14301: gap of 100 bp
* 14302	28801: contig of 14500 bp in length
* 28802	28901: gap of 100 bp
* 28902	40910: contig of 12009 bp in length
* 40911	41010: gap of 100 bp
* 41011	54579: contig of 13569 bp in length
* 54580	54679: gap of 100 bp
* 54680	62489: contig of 7810 bp in length
* 62490	62589: gap of 100 bp
* 62590	72710: contig of 10121 bp in length
* 72711	72810: gap of 100 bp
* 72811	81356: contig of 8546 bp in length
* 81357	81456: gap of 100 bp
* 81457	87558: contig of 6102 bp in length
* 87559	87658: gap of 100 bp
* 87659	95328: contig of 7670 bp in length
* 95329	95428: gap of 100 bp
* 95429	102558: contig of 7130 bp in length
* 102559	102658: gap of 100 bp
* 102659	107119: contig of 4461 bp in length
* 107120	107219: gap of 100 bp
* 107220	112453: contig of 5234 bp in length
* 112454	112553: gap of 100 bp
* 112554	117698: contig of 5145 bp in length
* 117699	117798: gap of 100 bp
* 117799	122290: contig of 4492 bp in length
* 122291	122390: gap of 100 bp
* 122391	126549: contig of 4159 bp in length
* 126550	126649: gap of 100 bp
* 126650	128718: contig of 2069 bp in length
* 128719	128818: gap of 100 bp
* 128819	132710: contig of 3892 bp in length
* 132711	132810: gap of 100 bp
* 132811	136147: contig of 3337 bp in length
* 136148	136247: gap of 100 bp
* 136248	139746: contig of 3499 bp in length
* 139747	139846: gap of 100 bp
* 139847	142112: contig of 2266 bp in length
* 142113	142212: gap of 100 bp
* 142213	144472: contig of 2260 bp in length
* 144473	144572: gap of 100 bp
* 144573	147200: contig of 2628 bp in length
* 147201	147300: gap of 100 bp
* 147301	150373: contig of 3073 bp in length
* 150374	150473: gap of 100 bp
* 150474	152540: contig of 2067 bp in length
* 152541	152640: gap of 100 bp
* 152641	155258: contig of 2618 bp in length
* 155259	155358: gap of 100 bp
* 155359	158220: contig of 2862 bp in length
* 158221	158320: gap of 100 bp
* 158321	160464: contig of 2144 bp in length
* 160465	160564: gap of 100 bp
* 160565	163294: contig of 2730 bp in length
* 163295	163394: gap of 100 bp

QY 1 GAGTGGTGGTGGTGACCGTGAAC 23  
|||||  
Db 158509 GAGTGGTGGTGGTGACAGTGAGC 158487



RESULT 5  
 AP002434/c  
 LOCUS AP002434 197834 bp DNA linear HTG 07-FEB-2001  
 DEFINITION Homo sapiens chromosome 18 clone RP11-716A12 map 18q12, WORKING  
 DRAFT SEQUENCE, 35 unordered pieces.  
 ACCESSION AP002434  
 VERSION AP002434.3 GI:12718825  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 197834)  
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 COMMENT On Feb 8, 2001 this sequence version replaced gi:12539453.  
 ----- Genome Center  
 Center: RIKEN Genomic Sciences Center(GSC)  
 Center code: RIKEN  
 Web site: http://hgp.gsc.riken.go.jp/  
 Contact: hattori@gsc.riken.go.jp  
 ----- Project Information  
 Center project name: HumDraft18  
 Center clone name: RP11-716A12  
 ----- Summary Statistics  
 Sequencing vector: PCR products; 100% of reads  
 Chemistry: Dye-terminator ET-amersham; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 180703 bases at least Q40  
 Consensus quality: 188444 bases at least Q30  
 Consensus quality: 192158 bases at least Q20  
 Insert size: 194434; sum-of-contigs  
 Quality coverage: 5.19x in Q20 bases; sum-of-contigs  
 -----  
 NOTE: This is a 'working draft' sequence. It currently consists of  
 35 contigs. The true order of the pieces is not known and their  
 order in this sequence record is arbitrary. Gaps between the  
 contigs are represented as runs N, but the exact sizes of the gaps  
 are unknown. This record will be updated with the finished sequence  
 as soon as it is available and the accession number will be  
 preserved  

1	20719 contig of	20719 bp in length
20820	38629 contig of	17810 bp in length
38730	54275 contig of	15546 bp in length
54376	68982 contig of	14607 bp in length
69083	79677 contig of	10595 bp in length
79778	88679 contig of	8902 bp in length
88780	98369 contig of	9590 bp in length
98470	106831 contig of	8362 bp in length
106932	114348 contig of	7417 bp in length
114449	121636 contig of	7188 bp in length

121737	130463	contig of	8727 bp in length
130564	138787	contig of	8224 bp in length
138888	146599	contig of	7712 bp in length
146700	153196	contig of	6497 bp in length
153297	157585	contig of	4289 bp in length
157686	161445	contig of	3760 bp in length
161546	166280	contig of	4735 bp in length
166381	169639	contig of	3259 bp in length
169740	172373	contig of	2634 bp in length
172474	175129	contig of	2656 bp in length
175230	177770	contig of	2541 bp in length
177871	178463	contig of	593 bp in length
178564	180496	contig of	1933 bp in length
180597	182311	contig of	1715 bp in length
182412	184451	contig of	2040 bp in length
184552	186011	contig of	1460 bp in length
186112	187285	contig of	1174 bp in length
187386	188988	contig of	1603 bp in length
189089	190330	contig of	1242 bp in length
190431	191545	contig of	1115 bp in length
191646	192770	contig of	1125 bp in length
192871	193982	contig of	1112 bp in length
194083	195402	contig of	1320 bp in length
195503	196583	contig of	1081 bp in length
196684	197834	contig of	1151 bp in length.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 35 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

*	1	20719: contig of 20719 bp in length
*	20720	20819: gap of 100 bp
*	20820	38629: contig of 17810 bp in length
*	38630	38729: gap of 100 bp
*	38730	54275: contig of 15546 bp in length
*	54276	54375: gap of 100 bp
*	54376	68982: contig of 14607 bp in length
*	68983	69082: gap of 100 bp
*	69083	79677: contig of 10595 bp in length
*	79678	79777: gap of 100 bp
*	79778	88679: contig of 8902 bp in length
*	88680	88779: gap of 100 bp
*	88780	98369: contig of 9590 bp in length
*	98370	98469: gap of 100 bp
*	98470	106831: contig of 8362 bp in length
*	106832	106931: gap of 100 bp
*	106932	114348: contig of 7417 bp in length
*	114349	114448: gap of 100 bp
*	114449	121636: contig of 7188 bp in length
*	121637	121736: gap of 100 bp
*	121737	130463: contig of 8727 bp in length
*	130464	130563: gap of 100 bp
*	130564	138787: contig of 8224 bp in length
*	138788	138887: gap of 100 bp

\* 138888 146599: contig of 7712 bp in length  
 \* 146600 146699: gap of 100 bp  
 \* 146700 153196: contig of 6497 bp in length  
 \* 153197 153296: gap of 100 bp  
 \* 153297 157585: contig of 4289 bp in length  
 \* 157586 157685: gap of 100 bp  
 \* 157686 161445: contig of 3760 bp in length  
 \* 161446 161545: gap of 100 bp  
 \* 161546 166280: contig of 4735 bp in length  
 \* 166281 166380: gap of 100 bp  
 \* 166381 169639: contig of 3259 bp in length  
 \* 169640 169739: gap of 100 bp  
 \* 169740 172373: contig of 2634 bp in length  
 \* 172374 172473: gap of 100 bp  
 \* 172474 175129: contig of 2656 bp in length  
 \* 175130 175229: gap of 100 bp  
 \* 175230 177770: contig of 2541 bp in length  
 \* 177771 177870: gap of 100 bp  
 \* 177871 178463: contig of 593 bp in length  
 \* 178464 178563: gap of 100 bp  
 \* 178564 180496: contig of 1933 bp in length  
 \* 180497 180596: gap of 100 bp  
 \* 180597 182311: contig of 1715 bp in length  
 \* 182312 182411: gap of 100 bp  
 \* 182412 184451: contig of 2040 bp in length  
 \* 184452 184551: gap of 100 bp  
 \* 184552 186011: contig of 1460 bp in length  
 \* 186012 186111: gap of 100 bp  
 \* 186112 187285: contig of 1174 bp in length  
 \* 187286 187385: gap of 100 bp  
 \* 187386 188988: contig of 1603 bp in length  
 \* 188989 189088: gap of 100 bp  
 \* 189089 190330: contig of 1242 bp in length  
 \* 190331 190430: gap of 100 bp  
 \* 190431 191545: contig of 1115 bp in length  
 \* 191546 191645: gap of 100 bp  
 \* 191646 192770: contig of 1125 bp in length  
 \* 192771 192870: gap of 100 bp  
 \* 192871 193982: contig of 1112 bp in length  
 \* 193983 194082: gap of 100 bp  
 \* 194083 195402: contig of 1320 bp in length  
 \* 195403 195502: gap of 100 bp  
 \* 195503 196583: contig of 1081 bp in length  
 \* 196584 196683: gap of 100 bp  
 \* 196684 197834: contig of 1151 bp in length.

FEATURES	Location/Qualifiers
source	1. .197834
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="18"
	/map="18q12"
	/clone="RP11-716A12"
misc_feature	1. .20719
	/note="assembly_fragment"
misc_feature	20820. .38629
	/note="assembly_fragment"

```

misc_feature      38730. .54275
                  /note="assembly_fragment"
misc_feature      54376. .68982
                  /note="assembly_fragment"
misc_feature      69083. .79677
                  /note="assembly_fragment"
misc_feature      79778. .88679
                  /note="assembly_fragment"
misc_feature      88780. .98369
                  /note="assembly_fragment"
misc_feature      98470. .106831
                  /note="assembly_fragment clone_end:T7 vector_side:left"
misc_feature      106932. .114348
                  /note="assembly_fragment"
misc_feature      114449. .121636
                  /note="assembly_fragment"
misc_feature      121737. .130463
                  /note="assembly_fragment"
misc_feature      130564. .138787
                  /note="assembly_fragment"
misc_feature      138888. .146599
                  /note="assembly_fragment"
misc_feature      146700. .153196
                  /note="assembly_fragment"
misc_feature      153297. .157585
                  /note="assembly_fragment"
misc_feature      157686. .161445
                  /note="assembly_fragment"
misc_feature      161546. .166280
                  /note="assembly_fragment"
misc_feature      166381. .169639

```

```

Query Match          86.1%;  Score 19.8;  DB 2;  Length 197834;
Best Local Similarity 91.3%;  Pred. No. 3.7e+02;
Matches    21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

```

```

Qy          1 GAGTGGTGGTGGTGACCGTGAAC 23
             |||||
Db          167230 GAGTGGTGGTGGTGACAGTGAGC 167208

```

# RESULT 6

AP001200/c

LOCUS AP001200 206736 bp DNA linear HTG 26-JUL-2000

DEFINITION Homo sapiens chromosome 18 clone RP11-807G14 map 18q12, WORKING  
DRAFT SEQUENCE, 21 unordered pieces.

ACCESSION AP001200

VERSION AP001200.3 GI:9501839

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 206736)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens 206,736 genomic DNA of 18q12

JOURNAL Published Only in DataBase (2000)  
REFERENCE 2 (bases 1 to 206736)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (18-FEB-2000) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan  
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-42-778-9923, Fax:81-42-778-9924)  
COMMENT On Jul 26, 2000 this sequence version replaced gi:8117646.

----- Genome Center  
Center: RIKEN Genomic Sciences Center(GSC)  
Center code: RIKEN  
Web site: http://hgp.gsc.riken.go.jp/  
Contact: hattori@gsc.riken.go.jp  
----- Project Information  
Center project name: HumDraft18  
Center clone name: RP11-807G14  
----- Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 200020 bases at least Q40  
Consensus quality: 202650 bases at least Q30  
Consensus quality: 203906 bases at least Q20  
Insert size: 204736; sum-of-contigs  
Quality coverage: 9.25x in Q20 bases; sum-of-contigs  
-----.

NOTE: This is a 'working draft' sequence. It currently consists of  
21 contigs. The true order of the pieces is not known and their  
order in this sequence record is arbitrary. Gaps between the  
contigs are represented as runs N, but the exact sizes of the gaps  
are unknown. This record will be updated with the finished sequence  
as soon as it is available and the accession number will be  
preserved

1	45999	contig of	45999	bp in length
46100	75940	contig of	29841	bp in length
76041	96923	contig of	20883	bp in length
97024	113708	contig of	16685	bp in length
113809	124498	contig of	10690	bp in length
124599	133447	contig of	8849	bp in length
133548	141890	contig of	8343	bp in length
141991	149359	contig of	7369	bp in length
149460	156631	contig of	7172	bp in length
156732	161520	contig of	4789	bp in length
161621	169375	contig of	7755	bp in length
169476	177104	contig of	7629	bp in length
177205	182218	contig of	5014	bp in length
182319	187278	contig of	4960	bp in length
187379	191932	contig of	4554	bp in length
192033	195629	contig of	3597	bp in length
195730	199121	contig of	3392	bp in length
199222	201939	contig of	2718	bp in length
202040	203612	contig of	1573	bp in length
203713	205170	contig of	1458	bp in length
205271	206736	contig of	1466	bp in length.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 21 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```
*      1      45999: contig of 45999 bp in length
*    46000    46099: gap of 100 bp
*    46100    75940: contig of 29841 bp in length
*    75941    76040: gap of 100 bp
*    76041    96923: contig of 20883 bp in length
*    96924    97023: gap of 100 bp
*    97024   113708: contig of 16685 bp in length
*   113709   113808: gap of 100 bp
*   113809   124498: contig of 10690 bp in length
*   124499   124598: gap of 100 bp
*   124599   133447: contig of 8849 bp in length
*   133448   133547: gap of 100 bp
*   133548   141890: contig of 8343 bp in length
*   141891   141990: gap of 100 bp
*   141991   149359: contig of 7369 bp in length
*   149360   149459: gap of 100 bp
*   149460   156631: contig of 7172 bp in length
*   156632   156731: gap of 100 bp
*   156732   161520: contig of 4789 bp in length
*   161521   161620: gap of 100 bp
*   161621   169375: contig of 7755 bp in length
*   169376   169475: gap of 100 bp
*   169476   177104: contig of 7629 bp in length
*   177105   177204: gap of 100 bp
*   177205   182218: contig of 5014 bp in length
*   182219   182318: gap of 100 bp
*   182319   187278: contig of 4960 bp in length
*   187279   187378: gap of 100 bp
*   187379   191932: contig of 4554 bp in length
*   191933   192032: gap of 100 bp
*   192033   195629: contig of 3597 bp in length
*   195630   195729: gap of 100 bp
*   195730   199121: contig of 3392 bp in length
*   199122   199221: gap of 100 bp
*   199222   201939: contig of 2718 bp in length
*   201940   202039: gap of 100 bp
*   202040   203612: contig of 1573 bp in length
*   203613   203712: gap of 100 bp
*   203713   205170: contig of 1458 bp in length
*   205171   205270: gap of 100 bp
*   205271   206736: contig of 1466 bp in length.
```

```
FEATURES                      Location/Qualifiers
  source                      1. .206736
                              /organism="Homo sapiens"
                              /mol_type="genomic DNA"
                              /db_xref="taxon:9606"
                              /chromosome="18"
                              /map="18q12"
                              /clone="RP11-807G14"
```

```

misc_feature      1. .45999
                  /note="assembly_fragment"
misc_feature      46100. .75940
                  /note="assembly_fragment"
misc_feature      76041. .96923
                  /note="assembly_fragment"
misc_feature      97024. .113708
                  /note="assembly_fragment"
misc_feature      113809. .124498
                  /note="assembly_fragment"
misc_feature      124599. .133447
                  /note="assembly_fragment clone_end:T7 vector_side:left"
misc_feature      133548. .141890
                  /note="assembly_fragment"
misc_feature      141991. .149359
                  /note="assembly_fragment"
misc_feature      149460. .156631
                  /note="assembly_fragment"
misc_feature      156732. .161520
                  /note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature      161621. .169375
                  /note="assembly_fragment"
misc_feature      169476. .177104
                  /note="assembly_fragment"
misc_feature      177205. .182218
                  /note="assembly_fragment"
misc_feature      182319. .187278
                  /note="assembly_fragment"
misc_feature      187379. .191932
                  /note="assembly_fragment"
misc_feature      192033. .195629
                  /note="assembly_fragment"
misc_feature      195730. .199121
                  /note="assembly_fragment"
misc_feature      199222. .201939
                  /note="assembly_fragment"
misc_feature      202040. .203612
                  /note="assembly_fragment"
misc_feature      203713. .205170
                  /note="assembly_fragment"
misc_feature      205271. .206736
                  /note="assembly_fragment"

```

# ORIGIN

```

Query Match          86.1%;  Score 19.8;  DB 2;  Length 206736;
Best Local Similarity 91.3%;  Pred. No. 3.7e+02;
Matches 21;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

```

```

QY          1 GAGTGGTGGTGGTGACCGTGAAC 23
             |||||
Db          138310 GAGTGGTGGTGGTGACAGTGAGC 138288

```

# RESULT 7

AC015563

LOCUS AC015563 214984 bp DNA linear PRI 31-JUL-2002

DEFINITION Homo sapiens chromosome 18, clone RP11-344B2, complete sequence.

ACCESSION AC015563  
 VERSION AC015563.11 GI:22024600  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 214984)  
 AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 18, clone RP11-344B2  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 214984)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lehoczký,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 214984)  
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUL-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 214984)  
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,



Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (31-JUL-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Jul 31, 2002 this sequence version replaced gi:21699684.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L1004  
Center clone name: 344\_B\_2  
-----

FEATURES	Location/Qualifiers
source	1. .214984 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="18" /map="18" /clone="RP11-344B2" /clone_lib="RPCI-11 Human Male BAC"
repeat_region	complement(683. .970) /rpt_family="AluSx"
repeat_region	complement(980. .1266) /rpt_family="AluSx"
repeat_region	1595. .1656 /rpt_family="MIR"
repeat_region	complement(1823. .1911) /rpt_family="AluJb"
unsure	complement(1904. .1908) /note="<30 qual SNGL region"
unsure	complement(1925. .2041) /note="<30 qual SNGL region"
repeat_region	complement(1935. .2003) /rpt_family="FRAM/FAM"
unsure	complement(1960) /note="probably G, possibly C"
unsure	complement(2070) /note="probably C, possibly A"
unsure	complement(2073. .2079)

	/note="<30 qual SNGL region"
unsure	complement(2084. .2088)
	/note="<30 qual SNGL region"
unsure	complement(2097. .2101)
	/note="<30 qual SNGL region"
repeat_region	2109. .2184
	/rpt_family="AluSp/q"
unsure	complement(2111. .2141)
	/note="<30 qual SNGL region"
repeat_region	complement(2574. .2653)
	/rpt_family="Charliel"
repeat_region	2654. .3007
	/rpt_family="MER7A"
repeat_region	complement(3008. .3047)
	/rpt_family="Charliel"
repeat_region	3048. .3181
	/rpt_family="FLAM_C"
repeat_region	complement(3182. .3477)
	/rpt_family="Charliel"
repeat_region	3478. .3781
	/rpt_family="AluJb"
repeat_region	complement(3782. .3832)
	/rpt_family="Charliel"
repeat_region	complement(3835. .3974)
	/rpt_family="AluSp/q"
repeat_region	complement(3975. .4559)
	/rpt_family="Charliel"
repeat_region	complement(4566. .4721)
	/rpt_family="AluSg/x"
repeat_region	complement(4743. .4933)
	/rpt_family="L1MC/D"
repeat_region	4994. .5147
	/rpt_family="L1MC/D"
repeat_region	complement(5227. .5616)
	/rpt_family="L1MC/D"
repeat_region	complement(5633. .6104)
	/rpt_family="MER41B"
repeat_region	6107. .6128
	/rpt_family="AT_rich"
repeat_region	complement(6137. .6177)
	/rpt_family="Alu"
repeat_region	complement(6183. .6491)
	/rpt_family="AluY"
repeat_region	6800. .7751
	/rpt_family="pTR5"
repeat_region	7756. .7853
	/rpt_family="LTR30"
repeat_region	7857. .8074
	/rpt_family="pTR5"
repeat_region	complement(8215. .8480)
	/rpt_family="L3"
repeat_region	complement(8516. .8846)
	/rpt_family="L3"
repeat_region	9053. .9189
	/rpt_family="AluJb"
repeat_region	9579. .9884
	/rpt_family="AluJb"

repeat\_region complement(9942. .10243)  
 /rpt\_family="AluSx"  
 repeat\_region complement(10260. .10603)  
 /rpt\_family="AluSq"  
 repeat\_region 11292. .12391  
 /rpt\_family="L2"  
 repeat\_region 12588. .12687  
 /rpt\_family="LTR16C"  
 repeat\_region 12725. .12881  
 /rpt\_family="(TTTC)n"  
 repeat\_region complement(12884. .13148)  
 /rpt\_family="AluJo"  
 repeat\_region 13155. .13250  
 /rpt\_family="CT-rich"  
 repeat\_region complement(13251. .13510)  
 /rpt\_family="AluJb"  
 repeat\_region 13512. .13602  
 /rpt\_family="LTR16C"  
 repeat\_region 13616. .13795  
 /rpt\_family="MER5A"  
 repeat\_region 14787. .14809  
 /rpt\_family="AT-rich"  
 repeat\_region complement(14888. .15002)  
 /rpt\_family="L2"  
 repeat\_region complement(15099. .15237)  
 /rpt\_family="MER5B"  
 repeat\_region complement(15318. .15405)

Query Match 86.1%; Score 19.8; DB 9; Length 214984;  
 Best Local Similarity 91.3%; Pred. No. 3.7e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
 |||||  
 Db 161619 GAGTGGTGGTGGTGACAGTGAGC 161641

# RESULT 8

AX655151/c

LOCUS AX655151 2000 bp DNA linear PAT 22-MAR-2003

DEFINITION Sequence 5021 from Patent WO03000898.

ACCESSION AX655151

VERSION AX655151.1 GI:29157965

KEYWORDS .

SOURCE Oryza sativa

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1

AUTHORS Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,  
 Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.

TITLE Plant genes involved in defense against pathogens

JOURNAL Patent: WO 03000898-A 5021 03-JAN-2003;

Syngenta Participations AG (CH)

FEATURES Location/Qualifiers

source 1. .2000

/organism="Oryza sativa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4530"

ORIGIN

Query Match 84.3%; Score 19.4; DB 6; Length 2000;  
Best Local Similarity 95.2%; Pred. No. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAA 22  
|||||  
Db 48 AGTGGTGGTGGTGACACTGAA 28

RESULT 9

AP003235/c

LOCUS AP003235 148892 bp DNA linear PLN 27-NOV-2003

DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,  
PAC clone:P0039A07.

ACCESSION AP003235 BA000010

VERSION AP003235.2 GI:13699092

KEYWORDS .

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1

AUTHORS Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T.,  
Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y.,  
Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K.,  
Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K.,  
Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y.,  
Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Itoh,S., Itoh,T.,  
Itoh,Y., Itoh,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S.,  
Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T.,  
Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M.,  
Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M.,  
Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T.,  
Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K.,  
Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,R., Yukawa,K.,  
Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y.,  
Yano,M., Jiang,J. and Gojobori,T.

TITLE The genome sequence and structure of rice chromosome 1

JOURNAL Nature 420 (6913), 312-316 (2002)

MEDLINE 22337376

PUBMED 12447438

REFERENCE 2 (bases 1 to 148892)

AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.

TITLE Direct Submission

JOURNAL Submitted (19-FEB-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT On Apr 19, 2001 this sequence version replaced gi:13027265.  
Genes were predicted from the integrated results of the following:

GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI NonRedundant Protein database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DDBJ accession no. and RGP clone ID.

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.

The orientation of the sequence is from SP6 to T7 of the PAC clone.

Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES	Location/Qualifiers
source	1. .148892 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="1" /clone="P0039A07"
gene	complement(4094. .5626) /gene="P0039A07.1"
CDS	complement(4094. .5626) /gene="P0039A07.1" /codon_start=1 /product="putative flavonol 3-O-glucosyltransferase" /protein_id="BAB64095.1" /db_xref="GI:15408689" /translation="MDKTIVLYPGLYVSHFVPMQLADALLEHGYAVALIHVTMDE DATFAAAVARVAAAAPSVTFHKLPRIHDPPIAITTIVGYLEMVRRYNERLREFLRSGV RGRSGGIAAVVVDAPSIEALDVARELGI PAYSFFASTASALAVFLHLPWFRARAASFE ELGDAPLIVPGVPPMPASHLMPELLEDPESETYRATVSMRLATLDADGILVNTFASLE PRAVGALGDPLFLPATGGGEPRRRVPPVYCVGPLVVGHHDDDERKENTRHECLAWLDE QPDRSVVFLCFGGTGAVTHSAEQMREIAAGLENSGHRFMWVVRAPRGGGDDLDALLPD GFLETRTSGHGLVVERWAPQADVLRHRSTGAFVTHCGWNSASEGITARVPMCLCWPLY AEQRMNKVFMVEEMGVGVEVAGWHWQRGELVMAEEIEGKIRLVMESEGERLRSSVAA HGEAAVAWRKDGAGAGSSRAALRRFLSDVGGRELRSVETLLLWAFHEIVVARIGLP LD"
gene	complement(join(7654. .7748,8483. .8657,9312. .9419, 9506. .9596,9722. .9847,10141. .10232,10383. .10444, 10536. .10665,10701. .10904)) /gene="P0039A07.2"
CDS	complement(join(7654. .7748,8483. .8657,9312. .9419, 9506. .9596,9722. .9847,10141. .10232,10383. .10444, 10536. .10665,10701. .10904)) /gene="P0039A07.2" /note="hypothetical protein similar to Arabidopsis thaliana K10A8_100" /codon_start=1 /protein_id="BAB64096.1"

```

/db_xref="GI:15408690"
/translation="MASKQMEEIQRKLAVLAYPRANAPAQSLLFAGVERYRLEWLFF
RYAATHATFPISASQSARLAAQLRSQLEIEGFVFSLLRRLLGDRSPFTQQNWQGDSLD
RDEENSRIQHLAEIANFLGITPSVDTEAIQGRGSYDERVELLCLIVDLVEASCYADNP
EWSVDEQLAKDVLVDSIAEKQAQIFSEECKLFPADVQIQSIYPLPDITELELKLSEY
TKKMSNLQLMVQELASKYDYNPNEDYAETELKLREHLQS FLETVKSFNMIYTKFLSNL
RSLRDSYAAMAAGSLASNEPSSVTKIISDCESALTFLNNSLSILSTSVAREQEFKTK
HFPVRSQPHDVSVDVTPPAYDCTS"
gene      join(11860. .12223,13217. .13494,14190. .14468)
          /gene="P0039A07.3"
CDS       join(11860. .12223,13217. .13494,14190. .14468)
          /gene="P0039A07.3"
          /note="contains EST CG1228_8A
          similar to Arabidopsis thaliana chromosome 1, F8K7.23
          unknown protein"
          /codon_start=1
          /protein_id="BAB64097.1"
          /db_xref="GI:15408691"
          /translation="MPASTAWAARLAEGGEGTESIVRCGSPLATTTTSPMPPPPHRG
          GGGGRDTSAFFAATLVLWAVSVGFEIGARGRELAPVAAGFAFFQAASAAVRAAVSRD
          PLFVNTAVSLLHSSLTSASVIFVLVNRWHNKDLKNMFEHEELFGGSWVGAYSALCFSC
          GYFAYDQDMLRYRLYSGRIPGILMHHLILLCFTLALYRNVNTINYLIITLVCKERKLR
          RMAGFRDYNRKIVKLEWVLNWTTFVSARVACHILITYKLIIDAHKFDSGIELPLALFG
          MAGMNLNIFLGLDLLKAYTLERNQQTHQD"
          gene      join(17702. .17740,17763. .17866,17895. .17931,18322.
.18474)
          /gene="P0039A07.4"
          CDS       join(17702. .17740,17763. .17866,17895. .17931,18322.
.18474)
          /gene="P0039A07.4"
          /note="hypothetical protein"
          /codon_start=1
          /protein_id="BAB64098.1"
          /db_xref="GI:15408692"
          /translation="MATTAPSAGHSPTDAPRQVRKPVVAPGKHRRPTPAACLETVVL
          DGPWCAGTFAAAGNGVFAADLQNP SGDDRAPCQRVAAIEGIRGDVAARAPVAHPILAA
          RPPANRHP"
          gene      join(18669. .18866,18983. .19099,20479. .20970)
          /gene="P0039A07.5"
          CDS       join(18669. .18866,18983. .19099,20479. .20970)
          /gene="P0039A07.5"
          /note="contains ESTs AU078383(S13149),AU078384(S13149)"
          /codon_start=1
          /product="putative photosystem II subunit (22KDa)
          precursor"
          /protein_id="BAB64099.1"
          /db_xref="GI:15408693"
          /translation="MAQSMLVSGANGTVAAASTSRLQPVVRPTPF SRLVLVLSQPSSSLGR
          AVSVKTVALFGRSKTKAAPARKAEPKPKFKTEDGIFGTSGGIGFTKENELFVGRVAML
          GFAASILGEAITGKGILAQLNLETGIP IYEAEPLLLFFILFTLLGAIGALGDRGSFVD
          DQPV TGLDKAVIAPGKGFRSALGLSEGGPLFGFTKANELFVGRLAQLGIAFSIIG EII
          TGKGALAQLNIETGVPINEIEPLVLFNVVFFFI AAINPGTGK FVSDDDEE"
          gene      complement(join(26718. .27104,27349. .27422,31544. .31689,
          31742. .31896))
          /gene="P0039A07.6"
          CDS       complement(join(26718. .27104,27349. .27422,31544. .31689,
          31742. .31896))

```

```

/gene="P0039A07.6"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64100.1"
/db_xref="GI:15408694"
/translation="MTPAARSSPSPAGGASAARGAGWVVMVARRLSVWGARAGGRGAG
EVGGGGSGGGRREASVVGVPPTREPSGRNWGCRVGPTVGVVRVAYRAGAGGGGGGPRG
FALKLRQLQLSTPTYLELELCQTRQRRTSVVGVGSGRRRVGGTAHGRTGGASGDDER
TAHGRRLAGMTRGRGQGGRSSAIAVELARPSPRRVQGGDWGRRRAGGDRAAARGRRK
GGDRAADGGWRRVAGEGRRVAGEGRRAGDGIGDFF"
gene
.33146,
33331. .33423,33516. .33608,35451. .35555,35657. .35755,
35859. .36200)
/gene="P0039A07.7"
join(31997. .32116,32271. .32345,32899. .33000,33093.
33331. .33423,33516. .33608,35451. .35555,35657. .35755,
35859. .36200)
CDS
.33146,
33331. .33423,33516. .33608,35451. .35555,35657. .35755,
35859. .36200)
/gene="P0039A07.7"
/note="contains ESTs C22684(S4284),C22468(C0086)"
/codon_start=1
/product="putative protein kinase SPK-3"
/protein_id="BAB64101.1"
/db_xref="GI:15408695"
/translation="MEKYEAVRDIGSGNFGVARLMRNRETRELVAVKCIERGHRI
DEN
VYREIINHRSRLRHPNIIRFKEVILTPTHLMIVMEFAAGGELFDRICDRGRFSEDEARY
FFQQLICGVSYCHHMQICHRDLKLENVLLDGSAPRLKICDFGYSKSSVLHSRPSKSAV
GTPAYIAPEVLSRREYDGLADVWSCGVTLVYMLVGAYPFEDQDDPKNIRKTIQRIMS
VQYKIPDYVHISAECKQLIARIFVNNPLRRITMKEIKSHPWFLKNLPRELTETAQAMY
YRRDNSVPSFSDQTSEEIMKIVQEARTMPKSSRTGYWSDAGSDEEEKEEEERPEENEE
EEDEYDKRVKEVHASGELRMSSLRI"
gene
complement(join(37460. .37678,37859. .38122,38204. .38296,
38688. .38882,38966. .39064,39185. .39433,39565. .39768,
39844. .39903,40032. .40214,40731. .41300,41402. .41470,
41576. .42154))
CDS
complement(join(37460. .37678,37859. .38122,38204. .38296,
38688. .38882,38966. .39064,39185. .39433,39565. .39768,
39844. .39903,40032. .40214,40731. .41300,41402. .41470,
41576. .42154))
/gene="P0039A07.8"
/note="hypothetical protein
similar to Arabidopsis thaliana chromosome 1, K9P8.10"
/codon_start=1
/protein_id="BAB64102.1"
/db_xref="GI:15408696"
/translation="MAESDGGGEASPSGGGGEGSPDPRRPPARPQLTKSRTISGSAAS
AFDRWGTNSSSSILVRRSSTAPLPPGAAPRGLLTVAVDEPSYAAPNGGAAMLDRDWC

```

```

Query Match          84.3%;  Score 19.4;  DB 8;  Length 148892;
Best Local Similarity 95.2%;  Pred. No. 5.6e+02;
Matches 20;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

```

```

Qy          2 AGTGGTGGTGGTGACCGTGAA 22
             ||||||||||||||| |||||
Db          79037 AGTGGTGGTGGTGACAGTGAA 79017

```

RESULT 10  
AP003566/c  
LOCUS AP003566 152736 bp DNA linear PLN 21-MAR-2002  
DEFINITION *Oryza sativa* (japonica cultivar-group) genomic DNA, chromosome 1,  
BAC clone:OSJNBb0008G24.  
ACCESSION AP003566  
VERSION AP003566.3 GI:19571107  
KEYWORDS .  
SOURCE *Oryza sativa* (japonica cultivar-group)  
ORGANISM *Oryza sativa* (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; *Oryza*.  
REFERENCE 1  
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.  
TITLE *Oryza sativa* (japonica cultivar-group) genomic DNA, chromosome 1,  
BAC clone:OSJNBb0008G24  
JOURNAL Published Only in Database (2001)  
REFERENCE 2 (bases 1 to 152736)  
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAY-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
COMMENT On Mar 21, 2002 this sequence version replaced gi:17026096.  
Genes were predicted from the integrated results of the following:  
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor  
(October 1998 version). The genomic sequence was searched against  
NCBI NonRedundant Protein database, nr  
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at  
RGP. Protein homologies of the coding regions were searched against  
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent  
the identified cDNA sequences using BLASTN 2.0 with the  
corresponding DDBJ accession no. and RGP clone ID.  
A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with EST homology (covering  
almost the entire length of partial sequence) is classified as an  
'unknown' protein. A gene predicted with a gene prediction program  
is classified as a 'hypothetical' protein.  
The orientation of the sequence is from -21M13 to M13rev of the BAC  
clone. This sequence of OSJNBb0008G24 clone has an overlap with  
P0039A07 clone(DDBJ: AP003235) at the position 1 to 79,839 of 5'  
end. The sequence of this clone starts at the position 69,054 of  
P0039A07. Detailed information on overlap and assembly quality  
together with annotation of this entry is available at  
http://rgp.dna.affrc.go.jp/GenomeSeq.html.  
FEATURES Location/Qualifiers  
source 1. .152736  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"



```

/db_xref="taxon:39947"
/chromosome="1"
/clone="OSJNBb0008G24"
gene      join(3191. .3194,3517. .3731)
          /gene="OSJNBb0008G24.1"
CDS       join(3191. .3194,3517. .3731)
          /gene="OSJNBb0008G24.1"
          /note="hypothetical protein"
          /codon_start=1
          /protein_id="BAB86532.1"
          /db_xref="GI:19571108"
          /translation="MAKQISQAGVDFDFVNNNDPSEREGGGEGAASPRSPATVNSLHL
PPASLAKVAARLSPKEMEKREKREKKE"
gene      join(5130. .5328,5437. .5444)
          /gene="OSJNBb0008G24.2"
CDS       join(5130. .5328,5437. .5444)
          /gene="OSJNBb0008G24.2"
          /note="hypothetical protein"
          /codon_start=1
          /protein_id="BAB86533.1"
          /db_xref="GI:19571109"
          /translation="MASSGAPPCSSALVPPCLFVIAMATLQSAVVFADAADTVAADRP
LSGSQRLVSSRGKFAIGFFQPD"
gene      complement(join(8162. .8585,8742. .9089,9151. .9324,
9455. .9530,9810. .9842,10302. .10926))
          /gene="OSJNBb0008G24.3"
CDS       complement(join(8162. .8585,8742. .9089,9151. .9324,
9455. .9530,9810. .9842,10302. .10926))
          /gene="OSJNBb0008G24.3"
          /note="hypothetical protein"
          /codon_start=1
          /protein_id="BAB86534.1"
          /db_xref="GI:19571110"
          /translation="MAARQRRRAGAARGQAHPRPRLSATVLADGGHACVPDSPRQRL
RPSATSPRASSPAAATAVLATILADGGYGRTRLCHRPRQRLCSRPSATFETSPRASSV
AAATAALATILGNLRDLPASVLTGGGRGHGRSRDLLLLRWRLCPRPPLRPSSPAAAPLR
NLPASVLADGGLVLSPLAFSLSSRPLAASWSCLILTARRRRVAFSPRRRWICHRKKYT
EGGGHRQPAMGDTTAVMAATGEAAAVGPTQRNGGDDTAAGDTASASSEASASVPPSS
TARRVRPPGRVKRALAAASKGVETVRGGERIEGHRDGKGSTALEKAGEAVVRKAGGT
ERGGAVPDDDDDDGDMFEEEDACSSSLAAASSDLLLES GHHHHRQLTAHPSMANPRI
AAALAATPAVVSSPSLSLFVFLTWICMPVAVDVSRWLGDRPGLHRSASSITDAAHP
AAADREGRRSWRSSSSSEARGSSFGFPRPSSASFSHLLGRSVPELAHAATPPPSLASS
SLPPQPPPPRHRLPTPSSPPAGRPAGLPSPERERERRKGETGSGRPADVAY"
gene      12551. .14755
          /gene="OSJNBb0008G24.4"
misc_feature 12551. .14755
          /gene="OSJNBb0008G24.4"
          /note="probably inactive due to stop codon(s) in CDS
pseudogene, similar to S-receptor kinase"
          /pseudo
gene      join(16120. .16385,16566. .16655,17390. .17576,18474.
.18563,
18768. .18890,19000. .19073,19979. .20000,20680. .20765,
20907. .20955,21398. .21454,22346. .22631,22777. .22802)
          /gene="OSJNBb0008G24.5"
CDS       join(16120. .16385,16566. .16655,17390. .17576,18474.
.18563,

```

18768. .18890,19000. .19073,19979. .20000,20680. .20765,  
20907. .20955,21398. .21454,22346. .22631,22777. .22802)  
/gene="OSJNBb0008G24.5"  
/note="contains ESTs AU086013(S11035),AU086012(S11035)  
unknown protein"  
/codon\_start=1  
/protein\_id="BAB86535.1"  
/db\_xref="GI:19571111"  
/translation="MEAVAGGGGGGGESVGELLRLAAAMVPAEHYALAAALAVSVLA  
YGFLELHFLGDLRLGRGGRVELTFHPASEIYHRVASKCRSLHGRYLATPWLASPHLQ  
TLFLGISGRPPSFITYKRMQIELQHPEVTGIAKKTEALNEDDVPVHNQNIKDLSPSHQQ  
NSETNMHLLLLCPYAHEIWKTYVKHMAYSMATKGCNTVVSNRHGLGGVSITSCLYNA  
GWTEDLREVINYLHHKYPKAPMLCVGTSIGANIVKFIHLATYTNETEMLLMPLFPDEE  
TSNILEKKVGDRLFISRKLVRQFYDKALAFGLKGYAKLHEPVLVRLANWEGIKKRSRIR  
EFDHHATCMVAKYEAQDHILRSSLESSIDKSPYVNVMEDGMIAPVTDDGPCDDITPSH  
QVNDIKQDNGDFTQQNEHTREVDDKNITEVNA MPSQSPEQSAGQQVEEHYVGKSSGII  
C"

gene complement(join(23130. .23140,23783. .23914,25110. .25247,  
26547. .26565))  
/gene="OSJNBb0008G24.6"

CDS complement(join(23130. .23140,23783. .23914,25110. .25247,  
26547. .26565))  
/gene="OSJNBb0008G24.6"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB86536.1"  
/db\_xref="GI:19571112"  
/translation="MGLGRAESHPKSKTSINQKLHFSLSGFAKSPVGSSRPPVADPHK  
TIKWQVLNVQITVGTFGTFGHICLLDQNNHIEFRSNLPSSVTNISIMLILGVVGP"

gene complement(28126. .28473)  
/gene="OSJNBb0008G24.7"

CDS complement(28126. .28473)  
/gene="OSJNBb0008G24.7"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB86537.1"  
/db\_xref="GI:19571113"  
/translation="MTKGLTQTGGKKMRKRGSDAIPRRDAAESAGRRTTKPRALLAVAG  
MDWDRIGGGHGDGEGGGGERRRRRARGNTSRVTVVANGETISYFFVGKNTNLRGP  
WEYEILGSGSSDF"

gene 29176. .30744  
/gene="OSJNBb0008G24.8"

CDS 29176. .30744  
/gene="OSJNBb0008G24.8"  
/note="contains ESTs  
C26205(C11841),D47784(S13471),AU092293(C11841),  
AU096321(S13471)"  
/codon\_start=1  
/product="putative zinc finger protein"  
/protein\_id="BAB86538.1"  
/db\_xref="GI:19571114"  
/translation="MDSGLGRSSETSLKALPSMASNATRNTDPDQQGVRFSSMDQPPC  
FARPGQSFPAPFPPLFGVQSSSLYLDDIEAKIGNQFESNPSPNNPTMDWDPQAMLSNL  
SFLEQKIKQVKDIVQSMSNRESQVAGGSSEAQAKQQLVTADLTCTIIQLISTAGSLP  
SMKNPISSNPALRHLSNTLCAPMILGTNCNLRPSANDEATIPDISKTHDYEELMNSLN  
TTQAESDEMNCQNPCGGEGSEPIPMEDHDVKESDDGGERENLPFGSYVVLQLEKEEI  
LAPHTHFCLICGKGFKRDANLRMHMRGHGDEYKTAALAKPSKDSSLESAPVTRYSCP

YVGCKRNKEHKKFQPLKTLICVKNHYKRSHCDKSYTCSRNTKKFSVIADLKTHEKHC  
GRDKWLCSCGTTFSRKDKLFQGHVALFQGHTPALPMDIKVTGASEQPQGSEAMNTMVG  
SAGYNFPGSSDDIPNLDMKMADDPYFSPLSFDPFCFGLDDFTRPGFDISENPFSL  
PSGSCSFGQNGDS"  
gene join(31857. .32254,32287. .32491,32897. .32990,33078.  
.33115,  
33618. .33866,33954. .34067,34258. .34341,34540. .34710,  
34792. .34905,34999. .35142)  
/gene="OSJNBb0008G24.9"  
CDS join(31857. .32254,32287. .32491,32897. .32990,33078.  
.33115,  
33618. .33866,33954. .34067,34258. .34341,34540. .34710,  
34792. .34905,34999. .35142)  
/gene="OSJNBb0008G24.9"  
/note="contains ESTs D24345 (R1764),D24345 (R1764) "  
/codon\_start=1  
/product="putative aspartate aminotransferase"  
/protein\_id="BAB86539.1"  
/db\_xref="GI:195711115"  
/translation="MRKNPSEQEEEEKETEKANEIEVQFPFRPTDPHLPQPQVDAQ  
VHHPCSIRTNTTTQTTYRCDTRSRAKRYHPPHGGHLRLHLLHSHQARLLLLLLLLLKS  
QLRLLRQVSIGVALGGWIFAGSLAAIGDVMASGRCRMASVVVRAEAVDATISPTVSA  
LRPSKTMAITDQATALRQAGVPVIGLAAGEPFDTPHVIAEAGMNAIKDGYTRYTPNA  
GTLELRKAICNKLQEENGISYSPDQVLVLI PAPYWVSYPEMATLAGATPVILPTSISE  
NFLLRPELLASKINEKSRLILCSPSNPTGSVYPKELLEIADIVKKYPRLLVLSDEI  
YEHIIYQPAKHTSFASLPGMWDRTLTVNGFSKAFAMTGWRLGYLAAPKHFVAACGKIQ  
SQFTSGASSISQKAGLALNLGYAGGEAVSTMVKAFQERRDYLVKSFKELPGVKISEP  
QGAFYLFIDFSSYYGSEVEGFGTIKDSESLCMFLEKAQVALVPGDAFGDDKCIRMSY  
AAALSTLQTAMEKIKEAVALIKPRVAAK"  
gene complement(join(35438. .35506,35602. .35670,36009. .36104,  
36139. .36337,37768. .37859,38719. .38745))  
/gene="OSJNBb0008G24.10"  
CDS complement(join(35438. .35506,35602. .35670,36009. .36104,  
36139. .36337,37768. .37859,38719. .38745))  
/gene="OSJNBb0008G24.10"  
/note="contains EST D41521 (S4063)  
unknown protein"

Query Match 84.3%; Score 19.4; DB 8; Length 152736;  
Best Local Similarity 95.2%; Pred. No. 5.6e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAA 22  
| | | | | | | | | | | | | | | |  
Db 9984 AGTGGTGGTGGTGACAGTGAA 9964

RESULT 11  
AY117122/c  
LOCUS AY117122 460 bp DNA linear PLN 31-MAY-2003  
DEFINITION Rhizopogon vesiculosus microsatellite locus Rve2.71.  
ACCESSION AY117122  
VERSION AY117122.1 GI:31281829  
KEYWORDS .  
SOURCE Rhizopogon vesiculosus  
ORGANISM Rhizopogon vesiculosus  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

Boletales; Suillineae; Rhizopogonaceae; Rhizopogon.

REFERENCE 1 (bases 1 to 460)

AUTHORS Kretzer,A.M., Dunham,S., Molina,R. and Spatafora,J.W.

TITLE Microsatellite markers reveal below ground clone structure in two species of Rhizopogon forming tuberculate ectomycorrhizae on Douglas-fir

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 460)

AUTHORS Kretzer,A.M., Dunham,S., Molina,R. and Spatafora,J.W.

TITLE Direct Submission

JOURNAL Submitted (03-JUN-2002) Environmental and Forest Biology, SUNY College of Environmental Science and Forestry, 350 Illick Hall, Syracuse, NY 13210-2788, USA

FEATURES Location/Qualifiers

source 1. .460

/organism="Rhizopogon vesiculosus"

/mol\_type="genomic DNA"

/strain="T20874"

/db\_xref="taxon:180088"

repeat\_region 1. .460

/note="microsatellite locus Rve2.71"

/rpt\_type=tandem

ORIGIN

Query Match 82.6%; Score 19; DB 8; Length 460;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGA 21

|||||||

Db 306 GTGGTGGTGGTGACCGTGA 288

RESULT 12

AB096080/c

LOCUS AB096080 2638 bp mRNA linear ROD 15-NOV-2002

DEFINITION Rattus norvegicus mRNA for soluble guanlate cyclase alpha 2 subunit, complete cds.

ACCESSION AB096080

VERSION AB096080.1 GI:25006392

KEYWORDS .

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1

AUTHORS Yao,Y., Yamamoto,T. and Suzuki,N.

TITLE rat soluble guanylate cyclase alpha 2 subunit

JOURNAL Published Only in Database (2002)

REFERENCE 2 (bases 1 to 2638)

AUTHORS Yao,Y., Yamamoto,T. and Suzuki,N.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2002) Yuko Yao, Graduate school of Science, Hokkaido University; N10W8 Kita-ku, Sapporo, Hokkaido 060-0810, Japan (E-mail:yyao@sci.hokudai.ac.jp, Tel:81-11-706-4459, Fax:81-11-706-4459)

```

FEATURES                      Location/Qualifiers
    source                    1. .2638
                               /organism="Rattus norvegicus"
                               /mol_type="mRNA"
                               /db_xref="taxon:10116"
    CDS                       421. .2613
                               /codon_start=1
                               /product="soluble guanylate cyclase alpha 2 subunit"
                               /protein_id="BAC24017.1"
                               /db_xref="GI:25006393"
                               /translation="MSRRKISSESFSSSLGSDYLETSPEEEGECPLSKLCWNGSRSPPG
PPGSRAAAMAATPVPAASVAAAAAAVAAGSKRAQRRRRVNLDLGEISISLLTAPSPQT
IHMTLKRTLQYYEHQVIGYRDAEKNFHNISNRCSSADHSNKEEIEDVSGILRCTANVL
GLKFQEIQERFGEEFFKICFDENERVLRAVGSTLQDFNFGFDALLEHIRTSTFGKQATL
ESPSFLCKELPEGTCLKLHYFHPHTVGFAMLGMIKAAGKRIYHLNVEVEQIENEKFC
DGSTPSNYSCLTFLIKECETTQITKNIPQGTSQIPTDLRISINTFCRTFPFHLMFDPN
MVLQLGEGLRKQLRCDNHKVLKFEDCFEIVSPKVNATFDRVLLRLSTPFVIRTKPEA
SGTDNEDKVMEIKGQMIHVPESNAILFLGSPCVDKLDDELIGRGLHLSDIPIDATRDV
ILVGEQAKAQDGLKKRMDKLKATLEKTHQALEEEKKKTVDLLYSIFPGDVAQQLWQRQ
QVQARKFDDVTMLFSDIVGFTAICAQCTPMQVISMLNELYTRFDHQCGFLDIYKVETI
GDAYCVASGLHRKSLCHAKPIALMALKMMELSEEVLTPDGRPIQMRIGIHSGLVLAGV
VGVMPRYCLFGNNVTLASKFESGSHPRRINISPTTYQLLKREDSFTFIPRSREELPD
NFPKEIPGVCYFLELRTGPKPPKPSLSSSRIKKVSYNIGTMFLRETSI"

```

# ORIGIN

```

Query Match          82.6%;  Score 19;  DB 10;  Length 2638;
Best Local Similarity 100.0%;  Pred. No. 1.2e+03;
Matches   19;  Conservative    0;  Mismatches    0;  Indels      0;  Gaps      0;

```

```

Qy      2 AGTGGTGGTGGTGACCGTG 20
        |||||||||||||||
Db      328 AGTGGTGGTGGTGACCGTG 310

```

## RESULT 13 AF109963/c

```

LOCUS      AF109963                2657 bp    mRNA    linear    ROD 04-DEC-2000
DEFINITION Rattus norvegicus soluble guanylyl cyclase alpha2 subunit (GUCY1A2)
            mRNA, complete cds.

```

```

ACCESSION  AF109963
VERSION    AF109963.2  GI:11528624
KEYWORDS   .

```

```

SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

```

```

REFERENCE  1 (bases 1 to 2657)
AUTHORS   Koglin,M. and Behrends,S.
TITLE     Cloning and functional expression of the rat alpha(2) subunit of
            soluble guanylyl cyclase
JOURNAL   Biochim. Biophys. Acta 1494 (3), 286-289 (2000)
MEDLINE   20571097
PUBMED    11121588
REFERENCE  2 (bases 2062 to 2350)
AUTHORS   Behrends,S.
TITLE     Direct Submission

```

JOURNAL Submitted (30-NOV-1998) Department of Pharmacology, University  
Hamburg, Martinistrasse 52, Hamburg 20251, Germany

REFERENCE 3 (bases 1 to 2657)

AUTHORS Behrends, S. and Koglin, M.

TITLE Direct Submission

JOURNAL Submitted (01-MAR-2000) Department of Pharmacology, University  
Hamburg, Martinistrasse 52, Hamburg 20251, Germany

REMARK Sequence update by submitter

COMMENT On Dec 4, 2000 this sequence version replaced gi:5381340.

FEATURES

source 1. .2657  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Wistar-Kyoto"  
/db\_xref="taxon:10116"  
/tissue\_type="aorta"

gene 1. .2657  
/gene="GUCY1A2"

CDS 421. .2613  
/gene="GUCY1A2"  
/EC\_number="4.6.1.2"  
/function="produces cGMP"  
/note="nitric oxide-sensitive"  
/codon\_start=1  
/product="soluble guanylyl cyclase alpha2 subunit"  
/protein\_id="AAD42949.2"  
/db\_xref="GI:11528625"  
/translation="MSRRKISSESFSSLGSDYLETSPREEGECPLSKLCWNGSRSPPG  
PPGSRAAAMAATPVPAASVAAAAA VAAGSKRAQRRRRVNLD SLGESISLLTAPSPQT  
IHMTLKRTLQYYEHQVIGYRDAEKNFHNISNRCSSADHSNKEEIEDVSGILRCTANVL  
GLKFQEIQERFGEEFFKICFDENERVLRAVGSTLQDFNFGFDALLEHIRT SFGKQATL  
ESPSFLCKELPEGT LKLHYFHPHTVGFAMLGMIKAAGKRIYHLNVEVEQIENEKFC  
DGSTPSNYSCLTFLIKECETTQITKNI PQGTSQIPTDLRISINTFCRTFPFHLMFDPN  
MVVLQLGEGLRKQLRCDNHKVLKFEDCFEIVSPKVNATFDRVLLRLSTPFVIRTKPEA  
SGTDNEDKVMEIKGQMIHVPE SNAILFLGSPCVDKLDLIGRGLHLSDIP IHDATRDV  
ILVGEQAKAQDGLKKRMDKLKATLEKTHQALEEEKKKTVDLLYSIFPGDVAQQLWQRQ  
QVQARKFDDVTMLFSDIVGFTAICAQCTPMQVISMLNELYTRFDHQCGFLDIYKVETI  
GDAYCVASGLHRKSLCHAKPIALMALKMMELSEEVLTPDGRPIQMRIGIHSGSVLAGV  
VGVRMPRYCLFGNNVT LASKFESGSHPRRINISPTTYQLLKREDSFTFI PRSREELPD  
NFPKEIPGVCYFLELRTGPKPPKPSLSSSRIKKVSYNIGTMFLRETSL"

# ORIGIN

Query Match 82.6%; Score 19; DB 10; Length 2657;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTG 20  
| | | | | | | | | | | | | | | | | |  
Db 328 AGTGGTGGTGGTGACCGTG 310

RESULT 14  
AC010480  
LOCUS AC010480 99995 bp DNA linear PRI 03-OCT-2001  
DEFINITION Homo sapiens chromosome 5 clone CTD-2315M5, complete sequence.  
ACCESSION AC010480  
VERSION AC010480.7 GI:15887281

KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 99995)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 99995)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 REFERENCE 3 (bases 1 to 99995)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 REFERENCE 4 (bases 1 to 99995)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 COMMENT On Oct 3, 2001 this sequence version replaced gi:10280744.  
 Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www-shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.6% of Sequence;  
 Estimated Total Number of Errors is 0.4.  
 STS Content:  
 WI-15758 G23282.  
 FEATURES Location/Qualifiers  
 source 1. .99995  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2315M5"  
 ORIGIN  
 Query Match 82.6%; Score 19; DB 9; Length 99995;  
 Best Local Similarity 100.0%; Pred. No. 8.6e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 3 GTGGTGGTGGTGACCGTGA 21  
 ||||||||||||||||  
 Db 62668 GTGGTGGTGGTGACCGTGA 62686  
 RESULT 15  
 AP004191/c  
 LOCUS AP004191 142325 bp DNA linear HTG 21-MAR-2002  
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone  
 OJ1524\_D08, \*\*\* SEQUENCING IN PROGRESS \*\*\*.  
 ACCESSION AP004191

VERSION AP004191.1 GI:15718448  
 KEYWORDS HTG; HTGS\_PHASE2.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.  
 REFERENCE 1  
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.  
 TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC  
 clone:OJ1524\_D08  
 JOURNAL Published Only in Database (2001)  
 REFERENCE 2 (bases 1 to 142325)  
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-SEP-2001) Takuji Sasaki, National Institute of  
 Agrobiological Sciences, Rice Genome Research Program; Kannondai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
 Tel:81-298-38-7441, Fax:81-298-38-7468)  
 COMMENT The nucleotide sequence of this BAC clone was generated by  
 combining Monsanto and RGP-Japan sequencing data.  
 NOTE: It currently consists of 1 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces is believed  
 to be correct as given, however the sizes of the gaps between them  
 are based on estimates that have provided by the submitter. This  
 sequence will be replaced by the finished sequence as soon as it is  
 available and the accession number will be preserved.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 FEATURES Location/Qualifiers  
 source 1. .142325  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="2"  
 /clone="OJ1524\_D08"  
 ORIGIN  
 Query Match 82.6%; Score 19; DB 2; Length 142325;  
 Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 3 GTGGTGGTGGTGACCGTGA 21  
 |||||  
 Db 99313 GTGGTGGTGGTGACCGTGA 99295

Search completed: March 11, 2004, 07:33:48  
 Job time : 1682 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 10, 2004, 15:54:07 ; Search time 264 Seconds  
(without alignments)  
370.108 Million cell updates/sec

Title: US-10-057-890A-26  
Perfect score: 23  
Sequence: 1 gagtgggtgggtgggtgaccgtgaac 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result			%		Query		DB	ID	Description
	No.	Score	Match	Length	Match	Length			
	1	23	100.0	23	6	ABS52921			Abs52921 Human CCR
c	2	23	100.0	82	6	ABS52916			Abs52916 Human CCR
	3	22	95.7	79	6	ABS52915			Abs52915 Human CCR
c	4	19.4	84.3	2000	7	ADA71696			Ada71696 Rice gene
c	5	19	82.6	2573	6	ABZ23782			Abz23782 Human mac
	6	18.4	80.0	370	4	AAK58413			Aak58413 Human imm
	7	18.4	80.0	2295	2	AAX82088			Aax82088 Human SIG

c	8	18.4	80.0	114793	4	AAD08215	Aad08215 Human gen
	9	18.2	79.1	593	5	AAS05438	Aas05438 Mammalian
	10	18.2	79.1	110000	4	AAI99682_12	Continuation (13 o
	11	18.2	79.1	110000	4	AAI99683_12	Continuation (13 o
	12	18	78.3	403	7	ACD95109	Acd95109 Human col
c	13	18	78.3	641	4	AAC86686	Aac86686 DNA encod
c	14	18	78.3	765	6	ABK90639	Abk90639 DNA encod
c	15	18	78.3	765	6	ABA05183	Aba05183 Murine pr
c	16	18	78.3	768	6	ABK90638	Abk90638 DNA encod
c	17	18	78.3	772	2	AAV48405	Aav48405 Dominant-
c	18	18	78.3	1320	7	ABZ21918	Abz21918 HisPrPDPC
c	19	18	78.3	1322	4	AAF86479	Aaf86479 Murine Pr
c	20	18	78.3	1322	6	AAD24907	Aad24907 Mouse PrP
	21	18	78.3	1742	2	AAV84449	Aav84449 Human sec
	22	18	78.3	1742	4	ABA83232	Aba83232 Human sec
	23	18	78.3	1742	8	ACH04733	Ach04733 Novel hum
	24	18	78.3	1742	8	ACD44543	Acd44543 Human cDN
c	25	17.8	77.4	494	7	ABZ54264	Abz54264 Aspergill
	26	17.8	77.4	553	9	ADC76828	Adc76828 DNA homol
	27	17.8	77.4	650	9	ADC75412	Adc75412 T harzian
c	28	17.8	77.4	1764	7	ADA70762	Ada70762 Rice gene
c	29	17.8	77.4	1764	9	ADB78932	Adb78932 Rice tran
c	30	17.8	77.4	2730	5	AAS87782	Aas87782 DNA encod
	31	17.4	75.7	55	2	AAX22313	Aax22313 HSV-1 DNA
	32	17.4	75.7	56	2	AAX22314	Aax22314 HSV-1 DNA
	33	17.4	75.7	56	2	AAX22312	Aax22312 HSV-1 DNA
	34	17.4	75.7	57	2	AAX22315	Aax22315 HSV-1 DNA
c	35	17.4	75.7	231	4	ABL22053	Abl22053 Drosophil
c	36	17.4	75.7	253	4	AAI23498	Aai23498 Probe #13
c	37	17.4	75.7	253	4	ABA68608	Aba68608 Human foe
c	38	17.4	75.7	253	4	AAI48814	Aai48814 Probe #17
c	39	17.4	75.7	253	4	ABA50651	Aba50651 Human bre
c	40	17.4	75.7	253	4	ABA35583	Aba35583 Probe #14
c	41	17.4	75.7	253	4	AAK42743	Aak42743 Human bon
c	42	17.4	75.7	253	4	AAK16968	Aak16968 Human bra
c	43	17.4	75.7	253	4	ABS42365	Abs42365 Human liv
c	44	17.4	75.7	253	5	AAI09123	Aai09123 Probe #91
c	45	17.4	75.7	253	6	ABS16793	Abs16793 Human gen

#### ALIGNMENTS

##### RESULT 1

ABS52921

ID ABS52921 standard; DNA; 23 BP.

XX

AC ABS52921;

XX

DT 15-NOV-2002 (first entry)

XX

DE Human CCR5-based scaffolded fusion protein oligonucleotide 68133.

XX

KW Scaffolded protein; CCR5; HIV; human immunodeficiency virus infection;

KW ECD; extracellular domain; metal chelating motif; zinc finger protein;

KW integral membrane protein; soluble loop; intracellular domain; ICD;

KW gene therapy; immunogen; viral infection; human; ds.

XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200260477-A1.  
XX  
PD 08-AUG-2002.  
XX  
PF 29-JAN-2002; 2002WO-US002377.  
XX  
PR 31-JAN-2001; 2001US-0265782P.  
PR 31-JAN-2001; 2001US-0265858P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Coleman TA, Mansfield B;  
XX  
DR WPI; 2002-643357/69.  
XX  
PT Novel scaffolded fusion polypeptide useful for therapeutic purposes or  
PT for screening molecules that bind/activate/inhibit/modulate the  
PT polypeptide, comprises a functional polypeptide domain fused to a  
PT scaffold domain.  
XX  
PS Example 1; Page 40; 64pp; English.  
XX  
CC The invention relates to a scaffolded fusion polypeptide comprising a  
CC functional polypeptide domain fused to a scaffold domain, where the  
CC functional polypeptide domain corresponds to a soluble loop of an  
CC integral membrane protein (e.g. human CCR5, a transmembrane receptor  
CC involved in HIV (human immunodeficiency virus) infection). Also included  
CC are; (1) a polypeptide comprising a scaffold domain; (2) a nucleic acid  
CC encoding the fusion polypeptide; (3) a vector cassette for the expression  
CC of the fusion polypeptide comprising an expression region operably linked  
CC to a promoter, where the expression region comprises a number of  
CC cassettes, each of which encodes a module, domain or strand of the fusion  
CC polypeptide and (4) a host cell comprising the vector or nucleic acid.  
CC The fusion polypeptide is useful for screening molecules that  
CC bind/activate/inhibit/modulate the fusion polypeptide, by expressing the  
CC fusion polypeptide from and identifying a molecule that binds to the  
CC fusion polypeptide. The fusion polypeptide is useful in diagnostic  
CC methods, in assays to identify compounds that interact with loops of  
CC fragments of an extracellular domain (ECD) or an intracellular domain  
CC (ICD) or to rapidly assay the function of mutated portions of mutant  
CC integral membrane proteins without having to produce significant  
CC quantities of the entire mutant integral membrane protein, to generate  
CC antibodies that recognise the integral membrane proteins from which they  
CC are designed, to competitively bind the ligand of a naturally occurring  
CC receptor in vitro or in vivo, to display and/or screen soluble domains  
CC from protein such as integral membrane proteins, to probe the structure  
CC of ECD or ICD, or both, of an integral protein membrane, to modulate the  
CC activity of a receptor in vivo, and for treating or preventing viral  
CC infection, preferably human HIV infection e.g. by gene therapy using the  
CC encoding nucleic acid. The present sequence is an oligonucleotide used to  
CC make a scaffolded protein based on the ECD region of human CCR5 (not  
CC defined)  
XX

SQ Sequence 23 BP; 4 A; 3 C; 11 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
|||||  
Db 1 GAGTGGTGGTGGTGACCGTGAAC 23

RESULT 2

ABS52916/c

ID ABS52916 standard; DNA; 82 BP.

XX

AC ABS52916;

XX

DT 15-NOV-2002 (first entry)

XX

DE Human CCR5-based scaffolded fusion protein oligonucleotide 66974.

XX

KW Scaffolded protein; CCR5; HIV; human immunodeficiency virus infection;

KW ECD; extracellular domain; metal chelating motif; zinc finger protein;

KW integral membrane protein; soluble loop; intracellular domain; ICD;

KW gene therapy; immunogen; viral infection; human; ds.

XX

OS Homo sapiens.

OS Synthetic.

XX

PN WO200260477-A1.

XX

PD 08-AUG-2002.

XX

PF 29-JAN-2002; 2002WO-US002377.

XX

PR 31-JAN-2001; 2001US-0265782P.

PR 31-JAN-2001; 2001US-0265858P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Coleman TA, Mansfield B;

XX

DR WPI; 2002-643357/69.

XX

PT Novel scaffolded fusion polypeptide useful for therapeutic purposes or

PT for screening molecules that bind/activate/inhibit/modulate the

PT polypeptide, comprises a functional polypeptide domain fused to a

PT scaffold domain.

XX

PS Example 1; Page 40; 64pp; English.

XX

CC The invention relates to a scaffolded fusion polypeptide comprising a

CC functional polypeptide domain fused to a scaffold domain, where the

CC functional polypeptide domain corresponds to a soluble loop of an

CC integral membrane protein (e.g. human CCR5, a transmembrane receptor

CC involved in HIV (human immunodeficiency virus) infection). Also included

CC are; (1) a polypeptide comprising a scaffold domain; (2) a nucleic acid

CC encoding the fusion polypeptide; (3) a vector cassette for the expression  
 CC of the fusion polypeptide comprising an expression region operably linked  
 CC to a promoter, where the expression region comprises a number of  
 CC cassettes, each of which encodes a module, domain or strand of the fusion  
 CC polypeptide and (4) a host cell comprising the vector or nucleic acid.  
 CC The fusion polypeptide is useful for screening molecules that  
 CC bind/activate/inhibit/modulate the fusion polypeptide, by expressing the  
 CC fusion polypeptide from and identifying a molecule that binds to the  
 CC fusion polypeptide. The fusion polypeptide is useful in diagnostic  
 CC methods, in assays to identify compounds that interact with loops of  
 CC fragments of an extracellular domain (ECD) or an intracellular domain  
 CC (ICD) or to rapidly assay the function of mutated portions of mutant  
 CC integral membrane proteins without having to produce significant  
 CC quantities of the entire mutant integral membrane protein, to generate  
 CC antibodies that recognise the integral membrane proteins from which they  
 CC are designed, to competitively bind the ligand of a naturally occurring  
 CC receptor in vitro or in vivo, to display and/or screen soluble domains  
 CC from protein such as integral membrane proteins, to probe the structure  
 CC of ECD or ICD, or both, of an integral protein membrane, to modulate the  
 CC activity of a receptor in vivo, and for treating or preventing viral  
 CC infection, preferably human HIV infection e.g. by gene therapy using the  
 CC encoding nucleic acid. The present sequence is an oligonucleotide used to  
 CC make a scaffolded protein based on the ECD region of human CCR5 (not  
 CC defined)

XX

SQ Sequence 82 BP; 19 A; 28 C; 15 G; 20 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 6; Length 82;

Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGGTGGTGGTGACCGTGAAC 23  
 ||||||||||||||||  
 Db 23 GAGTGGTGGTGGTGACCGTGAAC 1

# RESULT 3

ABS52915

ID ABS52915 standard; DNA; 79 BP.

XX

AC ABS52915;

XX

DT 15-NOV-2002 (first entry)

XX

DE Human CCR5-based scaffolded fusion protein oligonucleotide 66735.

XX

KW Scaffolded protein; CCR5; HIV; human immunodeficiency virus infection;  
 KW ECD; extracellular domain; metal chelating motif; zinc finger protein;  
 KW integral membrane protein; soluble loop; intracellular domain; ICD;  
 KW gene therapy; immunogen; viral infection; human; ds.

XX

OS Homo sapiens.

OS Synthetic.

XX

PN WO200260477-A1.

XX

PD 08-AUG-2002.

XX  
PF 29-JAN-2002; 2002WO-US002377.  
XX  
PR 31-JAN-2001; 2001US-0265782P.  
PR 31-JAN-2001; 2001US-0265858P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Coleman TA, Mansfield B;  
XX  
DR WPI; 2002-643357/69.  
XX  
PT Novel scaffolded fusion polypeptide useful for therapeutic purposes or  
PT for screening molecules that bind/activate/inhibit/modulate the  
PT polypeptide, comprises a functional polypeptide domain fused to a  
PT scaffold domain.  
XX  
PS Example 1; Page 40; 64pp; English.  
XX  
CC The invention relates to a scaffolded fusion polypeptide comprising a  
CC functional polypeptide domain fused to a scaffold domain, where the  
CC functional polypeptide domain corresponds to a soluble loop of an  
CC integral membrane protein (e.g. human CCR5, a transmembrane receptor  
CC involved in HIV (human immunodeficiency virus) infection). Also included  
CC are; (1) a polypeptide comprising a scaffold domain; (2) a nucleic acid  
CC encoding the fusion polypeptide; (3) a vector cassette for the expression  
CC of the fusion polypeptide comprising an expression region operably linked  
CC to a promoter, where the expression region comprises a number of  
CC cassettes, each of which encodes a module, domain or strand of the fusion  
CC polypeptide and (4) a host cell comprising the vector or nucleic acid.  
CC The fusion polypeptide is useful for screening molecules that  
CC bind/activate/inhibit/modulate the fusion polypeptide, by expressing the  
CC fusion polypeptide from and identifying a molecule that binds to the  
CC fusion polypeptide. The fusion polypeptide is useful in diagnostic  
CC methods, in assays to identify compounds that interact with loops of  
CC fragments of an extracellular domain (ECD) or an intracellular domain  
CC (ICD) or to rapidly assay the function of mutated portions of mutant  
CC integral membrane proteins without having to produce significant  
CC quantities of the entire mutant integral membrane protein, to generate  
CC antibodies that recognise the integral membrane proteins from which they  
CC are designed, to competitively bind the ligand of a naturally occurring  
CC receptor in vitro or in vivo, to display and/or screen soluble domains  
CC from protein such as integral membrane proteins, to probe the structure  
CC of ECD or ICD, or both, of an integral protein membrane, to modulate the  
CC activity of a receptor in vivo, and for treating or preventing viral  
CC infection, preferably human HIV infection e.g. by gene therapy using the  
CC encoding nucleic acid. The present sequence is an oligonucleotide used to  
CC make a scaffolded protein based on the ECD region of human CCR5 (not  
CC defined)  
XX  
SQ Sequence 79 BP; 15 A; 18 C; 26 G; 20 T; 0 U; 0 Other;

Query Match 95.7%; Score 22; DB 6; Length 79;  
Best Local Similarity 100.0%; Pred. No. 8.8;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAAC 23

|||||  
Db 18 AGTGGTGGTGGTGACCGTGAAC 39

RESULT 4

ADA71696/c

ID ADA71696 standard; DNA; 2000 BP.

XX

AC ADA71696;

XX

DT 20-NOV-2003 (first entry)

XX

DE Rice gene, SEQ ID 5021.

XX

KW Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.

XX

OS Oryza sativa.

XX

PN WO2003000898-A1.

XX

PD 03-JAN-2003.

XX

PF 22-JUN-2001; 2001WO-IB001105.

XX

PR 22-JUN-2001; 2001WO-IB001105.

XX

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX

DR WPI; 2003-175290/17.

XX

PT Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.

XX

PS Claim 27; SEQ ID NO 5021; 899pp; English.

XX

CC The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

XX

SQ Sequence 2000 BP; 432 A; 463 C; 554 G; 469 T; 0 U; 82 Other;

Query Match 84.3%; Score 19.4; DB 7; Length 2000;

Best Local Similarity 95.2%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAA 22  
 |||||  
 Db 48 AGTGGTGGTGGTGACAGTGAA 28

RESULT 5

ABZ23782/c

ID ABZ23782 standard; cDNA; 2573 BP.

XX

AC ABZ23782;

XX

DT 20-JUN-2003 (first entry)

XX

DE Human macroprotein 1022-17.6 cDNA.

XX

KW Human; macroprotein; 1022-17.6; digestive ulcer; diabetes; gene; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 669. .1151

FT /\*tag= a

FT /product= "macroprotein 1022-17.6"

XX

PN CN1355214-A.

XX

PD 26-JUN-2002.

XX

PF 24-NOV-2000; 2000CN-00127559.

XX

PR 24-NOV-2000; 2000CN-00127559.

XX

PA (UYFU-) UNIV FUDAN.

XX

PI Mao Y, Xie Y;

XX

DR WPI; 2002-751456/82.

DR P-PSDB; ABP60147.

XX

PT A human macroprotein 1022-17.6 polypeptide and polynucleotide for  
 PT encoding it.

XX

PS Claim 6; Page 26-27 (disclosure); 34pp; Chinese.

XX

CC The invention relates to a human macroprotein 1022-17.6 polypeptide. Also  
 CC disclosed are the polynucleotide encoding the polypeptide, and a method  
 CC for preparing the polypeptide using DNA recombination techniques. The  
 CC polypeptide is used for treating diseases including digestive ulcers and  
 CC diabetes. The current sequence represents the human macroprotein 1022-  
 CC 17.6 encoding cDNA

XX

SQ Sequence 2573 BP; 956 A; 630 C; 499 G; 488 T; 0 U; 0 Other;

Query Match 82.6%; Score 19; DB 6; Length 2573;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Qy 3 GTGGTGGTGGTGACCGTGA 21  
|||||||  
Db 838 GTGGTGGTGGTGACCGTGA 820

RESULT 6

AAK58413

ID AAK58413 standard; cDNA; 370 BP.

XX

AC AAK58413;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:3473.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ss.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US001354.

XX

PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-483426/52.  
DR P-PSDB; AAM85632.  
XX

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
PS Claim 1; SEQ ID NO 3473; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 370 BP; 113 A; 68 C; 68 G; 113 T; 0 U; 8 Other;

Query Match 80.0%; Score 18.4; DB 4; Length 370;  
Best Local Similarity 86.4%; Pred. No. 3e+02;  
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAA 22  
|||||||:| | | |  
Db 244 GAGTGGTGGTKGTTACCTTGAA 265

# RESULT 7

AAX82088

ID AAX82088 standard; DNA; 2295 BP.

XX

AC AAX82088;

XX

DT 20-SEP-1999 (first entry)

XX

DE Human SIGP encoding DNA (clone ID 2965248).

XX

KW Signal-peptide containing protein; SIGP; human; cancer; immune response;  
KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma; AIDS;  
KW Addison's disease; adult respiratory distress syndrome; allergy; anemia;  
KW asthma; atherosclerosis; bronchitis; cholecystitis; Crohn's disease;  
KW ulcerative colitis; atopic dermatitis; dermatomyositis; emphysema;  
KW diabetes mellitus; atrophic gastritis; glomerulonephritis; gout; trauma;  
KW Grave's Disease; hypereosinophilia; irritable bowel syndrome; infection;  
KW lupus erythematosus; multiple sclerosis; myasthenia gravis; inflammation;  
KW osteoarthritis; osteoporosis; pancreatitis; polymyositis; scleroderma;  
KW rheumatoid arthritis; Sjogren's syndrome; autoimmune thyroiditis; ss.

XX

OS Homo sapiens.

XX

PN WO9933981-A2.  
 XX  
 PD 08-JUL-1999.  
 XX  
 PF 22-DEC-1998; 98WO-US027598.  
 XX  
 PR 31-DEC-1997; 97US-00002485.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Lal P, Hillman JL, Corley NC, Guegler KJ, Baughn MR, Sather SK;  
 PI Shah P;  
 XX  
 DR WPI; 1999-430242/36.  
 DR P-PSDB; AAY21853.  
 XX  
 PT Human signal-peptide containing protein coding sequences used to treat  
 PT cancer and immune responses.  
 XX  
 PS Claim 9; Page 97; 99pp; English.  
 XX  
 CC The invention provides human signal-peptide containing proteins (SIGP)  
 CC (AAY21841-855) and polynucleotides (AAX82076-90) encoding the proteins. A  
 CC host cell containing a vector comprising SIGP DNA can be used to produce  
 CC the SIGP protein. The SIGP protein can be used, in conjuncture with a  
 CC pharmaceutical carrier to treat or prevent a cancer. An antagonist of the  
 CC SIGP protein can be used to treat or prevent a cancer or an immune  
 CC response. The cancers that can be treated or prevented include sarcomas,  
 CC adenocarcinomas, leukemia's, lymphomas, melanomas, teratocarcinomas,  
 CC myelomas and cancers of the adrenal gland, bladder, bone, bone marrow,  
 CC brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract,  
 CC heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis,  
 CC prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and  
 CC uterus. The immune responses that can be treated or prevented include,  
 CC AIDS, Addison's disease, adult respiratory distress syndrome, allergies,  
 CC anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's  
 CC disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes  
 CC mellitus, emphysema, atrophic gastritis, glomerulonephritis, Grave's  
 CC disease, gout, hypereosinophilia, irritable bowel syndrome, lupus  
 CC erythematosus, multiple sclerosis, myasthenia gravis, myocardial or  
 CC pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,  
 CC polymyositis, rheumatoid arthritis, scleroderma, Sjogren's syndrome, and  
 CC autoimmune thyroiditis, complications of cancer, infections, and trauma  
 XX  
 SQ Sequence 2295 BP; 451 A; 707 C; 735 G; 402 T; 0 U; 0 Other;  
  
 Query Match 80.0%; Score 18.4; DB 2; Length 2295;  
 Best Local Similarity 95.0%; Pred. No. 3.5e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 GAGTGGTGGTGGTGACCGTG 20  
 ||| |||||  
 Db 877 GAGAGGTGGTGGTGACCGTG 896

RESULT 8  
 AAD08215/c

ID AAD08215 standard; DNA; 114793 BP.  
XX  
AC AAD08215;  
XX  
DT 08-AUG-2001 (first entry)  
XX  
DE Human genome from BAC clone, hbml68.  
XX  
KW Human; DNA helicase; NHL; cytostatic; neoplastic disorder; xeroderma;  
KW genetic disorder; multiple sclerosis; pigmentosum; Cockayne's syndrome;  
KW Bloom's syndrome; Werner's syndrome; therapy; chromosome 20; M68/DcR3;  
KW SCLIP; ARP; BAC clone hbml68; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 47095..85316  
FT /\*tag= a  
FT /note= "Corresponds to human DNA helicase, NHL gene"  
FT CDS 48688..84857  
FT /\*tag= b  
FT /product= "Human DNA helicase, NHL"  
XX  
PN WO200142434-A1.  
XX  
PD 14-JUN-2001.  
XX  
PF 07-DEC-2000; 2000WO-US033065.  
XX  
PR 09-DEC-1999; 99US-0169970P.  
XX  
PA (MERI ) MERCK & CO INC.  
XX  
PI Liu X, Bai C, Metzker ML;  
XX  
DR WPI; 2001-381666/40.  
DR P-PSDB; AAE03801.  
XX  
PT Novel polynucleotide encoding mammalian DNA helicase, NHL, useful for  
PT screening and measuring levels of NHL, and for formulating kits suitable  
PT for detecting and typing NHL.  
XX  
PS Claim 23; Page 21-78; 169pp; English.  
XX  
CC The invention relates to human DNA helicase protein, NHL and its  
CC corresponding DNA molecule. NHL gene is localised on human chromosome 20  
CC (20q13.3). NHL protein and its DNA are useful for treating various  
CC neoplastic disorders and genetic disorders such as multiple sclerosis,  
CC including xeroderma, pigmentosum, Cockayne's syndrome, Bloom's syndrome  
CC and Werner's syndrome. NHL protein is useful for selecting compounds  
CC active against neoplastic disorders. NHL protein is useful for screening  
CC and measuring levels of NHL, and for formulating kits suitable for  
CC detecting and typing NHL. The invention also relates to a method for  
CC identifying modulators of NHL activity. The present DNA sequence is human  
CC genome from BAC clone, hbml68. This human genomic DNA contains M68/DcR3,  
CC NHL DNA helicase, SCLIP and ARP gene at chromosome location 20 (20q13.3)  
XX

SQ Sequence 114793 BP; 24123 A; 32916 C; 31886 G; 25868 T; 0 U; 0 Other;

Query Match 80.0%; Score 18.4; DB 4; Length 114793;  
Best Local Similarity 95.0%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGTGGTGGTGGTGACCGTG 20  
||| |||||  
Db 30238 GAGAGGTGGTGGTGACCGTG 30219

RESULT 9

AAS05438

ID AAS05438 standard; DNA; 593 BP.

XX

AC AAS05438;

XX

DT 07-SEP-2001 (first entry)

XX

DE Mammalian vestibular system geotactic behaviour modulator gene #38.

XX

KW Mammalian vestibular system; invertebrate; geotactic behaviour; vertigo;

KW graviperceptive disorder; motion sickness; labyrinthitis; syphilis; ds;

KW Meniere's disease; acoustic neuroma; multiple sclerosis; epilepsy;

KW trauma; infection of the middle ear; ototoxic agent exposure.

XX

OS Drosophila melanogaster.

XX

PN WO200140519-A2.

XX

PD 07-JUN-2001.

XX

PF 01-DEC-2000; 2000WO-US032639.

XX

PR 02-DEC-1999; 99US-0168579P.

PR 26-SEP-2000; 2000US-00669751.

XX

PA (NEUR-) NEUROSCIENCES RES FOUND INC.

XX

PI Greenspan RJ;

XX

DR WPI; 2001-356159/37.

XX

PT New isolated nucleic acid having mammalian vestibular system-modulating

PT activity useful in the treatment of disorders such as motion sickness and

PT vertigo.

XX

PS Claim 59; Page 104; 179pp; English.

XX

CC The sequences shown in AAS05401-AAS05661 represent DNA with mammalian  
CC vestibular system-modulating activity. The DNA sequences can be used in a  
CC method whereby a first and second strain of an invertebrate is obtained,  
CC and both are subjected to conditions in which the strains exhibit  
CC different geotactic behaviour. Genes that are differentially expressed in  
CC the first strain relative to the second strain are then identified.  
CC Mammalian genes having substantially the same nucleic acid sequence as  
CC these modulate the mammalian vestibular system. Compounds containing

CC these genes are used to decrease the symptoms of graviperceptive  
 CC disorders such as motion sickness, vertigo, labyrinthitis, Meniere's  
 CC disease, acoustic neuroma, multiple sclerosis, syphilis, trauma,  
 CC infection of the middle ear, exposure to ototoxic agents and epilepsy  
 XX  
 SQ Sequence 593 BP; 129 A; 153 C; 139 G; 172 T; 0 U; 0 Other;

Query Match 79.1%; Score 18.2; DB 5; Length 593;  
 Best Local Similarity 87.0%; Pred. No. 3.8e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
 |||||  
 Db 87 GAGTGGTGGTGGTGACCCCGACC 109

# RESULT 10

AAI99682\_12

Continuation (13 of 45) of AAI99682 from base 1200001 (Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1. )

WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000



WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 79.1%; Score 18.2; DB 4; Length 110000;  
 Best Local Similarity 87.0%; Pred. No. 5.7e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
 ||||| ||||| |||||  
 Db 84960 GAGTGGTTGTGGTGACCGGCAC 84982

# RESULT 11

AAI99683\_12

Continuation (13 of 44) of AAI99683 from base 1200001 (Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2. )

WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683

WP	Fragment Name	Begin	End
WP	AAI99683_00	1	110000
WP	AAI99683_01	100001	210000
WP	AAI99683_02	200001	310000
WP	AAI99683_03	300001	410000
WP	AAI99683_04	400001	510000
WP	AAI99683_05	500001	610000
WP	AAI99683_06	600001	710000
WP	AAI99683_07	700001	810000
WP	AAI99683_08	800001	910000
WP	AAI99683_09	900001	1010000
WP	AAI99683_10	1000001	1110000
WP	AAI99683_11	1100001	1210000
WP	AAI99683_12	1200001	1310000
WP	AAI99683_13	1300001	1410000
WP	AAI99683_14	1400001	1510000
WP	AAI99683_15	1500001	1610000
WP	AAI99683_16	1600001	1710000
WP	AAI99683_17	1700001	1810000
WP	AAI99683_18	1800001	1910000
WP	AAI99683_19	1900001	2010000
WP	AAI99683_20	2000001	2110000
WP	AAI99683_21	2100001	2210000
WP	AAI99683_22	2200001	2310000
WP	AAI99683_23	2300001	2410000
WP	AAI99683_24	2400001	2510000
WP	AAI99683_25	2500001	2610000
WP	AAI99683_26	2600001	2710000
WP	AAI99683_27	2700001	2810000
WP	AAI99683_28	2800001	2910000
WP	AAI99683_29	2900001	3010000
WP	AAI99683_30	3000001	3110000

WP	AAI99683_31	3100001	3210000
WP	AAI99683_32	3200001	3310000
WP	AAI99683_33	3300001	3410000
WP	AAI99683_34	3400001	3510000
WP	AAI99683_35	3500001	3610000
WP	AAI99683_36	3600001	3710000
WP	AAI99683_37	3700001	3810000
WP	AAI99683_38	3800001	3910000
WP	AAI99683_39	3900001	4010000
WP	AAI99683_40	4000001	4110000
WP	AAI99683_41	4100001	4210000
WP	AAI99683_42	4200001	4310000
WP	AAI99683_43	4300001	4403765

Query Match 79.1%; Score 18.2; DB 4; Length 110000;  
 Best Local Similarity 87.0%; Pred. No. 5.7e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
 ||||| ||||| | ||  
 Db 84429 GAGTGGTTGTGGTGACCGGCAC 84451

# RESULT 12

ACD95109

ID ACD95109 standard; cDNA; 403 BP.

XX

AC ACD95109;

XX

DT 23-SEP-2003 (first entry)

XX

DE Human colon cancer cell expressed cDNA #3521.

XX

KW Open reading frame detection; genome sequencing; colon cancer;

KW breast cancer; population genome analysis; genetic shift; cancer;

KW antibiotic resistance; antibiotic non-tolerance; congenital disease;

KW agriculture; food crop genome; resistance gene; retrovirus;

KW influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;

KW gene; ss.

XX

OS Homo sapiens.

XX

PN US2002155438-A1.

XX

PD 24-OCT-2002.

XX

PF 27-SEP-1999; 99US-00406117.

XX

PR 20-NOV-1998; 98US-00196716.

XX

PA (SIMP/) SIMPSON A J G.

PA (NETO/) NETO E D.

PA (BREN/) BRENTANI R R.

XX

PI Simpson AJG, Neto ED, Brentani RR;

XX

DR WPI; 2003-182626/18.

XX  
PT Determining open reading frames of genome of an organism e.g. a human  
PT suffering from cancer involves use of single oligonucleotide primer at  
PT low stringency for preparing single-stranded cDNA from mRNA of  
PT individual.

XX  
PS Example 9; Page 515-516; 959pp; English.

XX  
CC The invention describes a method of determining open reading frames in  
CC the genome of organism, comprising contacting mRNA from cell of organism  
CC with a single oligonucleotide primer (I) at low stringency, preparing  
CC single-stranded cDNA by reverse transcribing mRNA with (I), amplifying  
CC cDNA, sequencing the product, and repeating the contacting, preparing  
CC and amplifying steps with different primers and sequencing resulting  
CC nucleic acids. The method is useful for: determining that a known  
CC nucleotide sequence from a genome of an organism corresponds to a  
CC nucleotide sequence of an open reading frame; for preparing a contig,  
CC nucleic acid molecule from a genome of an organism; and for sequencing  
CC all or part of a genome of an organism. mRNA is obtained from mammalian  
CC or human cell which is associated with a pathological condition e.g. a  
CC colon cancer or breast cancer cell. The method is useful for analyses of  
CC populations of subjects and can be used to carry out genetic analyses of  
CC large or small populations. further, it can be used to study living  
CC systems to determine if, e.g. there have been genetic shifts which render  
CC an individual or population more or less likely to be afflicted with  
CC diseases such as cancer, to determine antibiotic resistance or non-  
CC tolerance, and so forth. The method can also be used in the study of  
CC congenital diseases, and the risk of affliction to a foetus, as well as  
CC the study of whether the conditions are likely to be passed to offspring  
CC through ova or sperm. The analyses for pathological conditions can be  
CC carried out in all animals, plants, birds, fish, etc. Using this method,  
CC in the area of agriculture, for example the genomes of food crops can be  
CC studied to determine if resistance genes are present, defects in plant  
CC genomes can also be studied in this way. Similarly, the method permits  
CC determination of the pathogens which integrate into the genome, such as  
CC retroviruses and other integrating viruses such as influenza virus, have  
CC undergone shifts or mutations, which may require different approaches to  
CC therapy. This method is also applied to eukaryotic pathogens, such as  
CC trypanosomes, different types of Plasmodium, etc. The method essentially  
CC eliminates sequencing of non-coding portions. This sequence represents a  
CC polynucleotide isolated from human colon cancer cell cDNA library

XX  
SQ Sequence 403 BP; 57 A; 126 C; 114 G; 106 T; 0 U; 0 Other;

Query Match 78.3%; Score 18; DB 7; Length 403;  
Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GTGGTGGTGGTGACCGTG 20  
| | | | | | | | | | | | | | | | | |  
Db 57 GTGGTGGTGGTGACCGTG 74

RESULT 13  
AAC86686/c  
ID AAC86686 standard; DNA; 641 BP.  
XX

AC AAC86686;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE DNA encoding a mouse prion protein.  
 XX  
 KW SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;  
 KW aggregation; fibril; phenotypic alteration; gene therapy;  
 KW disease resistance; plant pigmentation; prion disease; ss.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1. .636  
 FT /\*tag= a  
 FT /product= "prion protein"  
 XX  
 PN WO200075324-A2.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 09-JUN-2000; 2000WO-US015876.  
 XX  
 PR 09-JUN-1999; 99US-0138833P.  
 XX  
 PA (ARCH-) ARCH DEV CORP.  
 XX  
 PI Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;  
 XX  
 DR WPI; 2001-061723/07.  
 DR P-PSDB; AAB30801.  
 XX  
 PT New nucleic acid encoding chimeric proteins with self-assembly  
 PT properties, useful e.g. for diagnosis and treatment of prion diseases,  
 PT also related aggregates, fibrils and polymers.  
 XX  
 PS Example 4; Page 136-137; 188pp; English.  
 XX  
 CC The present sequence encodes a prion protein. The specification describes  
 CC chimeric polypeptides, which comprise at least one SCHAG (self-coalesces  
 CC into higher-order aggregates) amino acid sequence fused in frame with a  
 CC polypeptide of interest (which is other than a marker protein, a  
 CC glutathione-S-transferase or a staphylococcal nuclear protein). The  
 CC specification also describes chimeric polypeptides that comprises an  
 CC amyloidogenic domain that causes aggregation into fibrils. The chimeric  
 CC polypeptides are used to prepare polymers with multiple reactivities,  
 CC e.g. derivatised with enzymes, or specific binding partners, and useful  
 CC e.g. for performing multi-step chemical reactions. They can be used  
 CC create an inducible, or stable phenotypic alteration in a cell, e.g. for  
 CC gene therapy, protein production, imparting disease resistance to plants,  
 CC altering plant pigmentation and for diagnosis and treatment of prion  
 CC diseases  
 XX  
 SQ Sequence 641 BP; 164 A; 173 C; 209 G; 95 T; 0 U; 0 Other;

Query Match 78.3%; Score 18; DB 4; Length 641;  
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTG 20  
|||||||  
Db 515 GTGGTGGTGGTGACCGTG 498

RESULT 14

ABK90639/c

ID ABK90639 standard; DNA; 765 BP.

XX

AC ABK90639;

XX

DT 05-NOV-2002 (first entry)

XX

DE DNA encoding mouse prion protein related peptide.

XX

KW Prion; mouse; gene; ds; follicular dendritic cells; FDC; infection;

KW blood preparation; food; cosmetic; CJD; Creutzfeldt-Jacob disease.

XX

OS Mus musculus.

XX

FH Key Location/Qualifiers

FT CDS 1. .766

FT /\*tag= a

FT /product= "Prion related protein"

XX

PN WO200261418-A1.

XX

PD 08-AUG-2002.

XX

PF 31-JAN-2002; 2002WO-JP000803.

XX

PR 31-JAN-2001; 2001JP-00024279.

XX

PA (TOHO ) UNIV TOHOKU.

XX

PI Kitamoto T, Miyoshi K, Mohri S;

XX

DR WPI; 2002-619277/66.

DR P-PSDB; ABG31906.

XX

PT Screening (non-)human prion disease infection factor based on abnormal

PT prion protein sedimentation in non-human follicular dendritic cells as

PT indication, applicable in safety test on e.g. drugs and cosmetics.

XX

PS Disclosure; Page 59-61; 69pp; Japanese.

XX

CC This invention relates to a novel method for screening human or non-  
CC human prion disease infection factor in a sample by using abnormal prion  
CC protein sedimentation in non-human follicular dendritic cells (FDC) as  
CC indication. The method of the invention is useful for screening (non-)  
CC human prion disease infection factor, which is applicable in safety tests  
CC on drugs like blood preparations, foods and cosmetics, and for developing  
CC drugs for e.g. CJD, as well as for early diagnosis of Creutzfeldt-Jacob  
CC disease (CJD). The method of the invention is simple and quick. The  
CC present sequence represents the DNA sequence encoding the mouse prion

CC related protein of the invention

XX

SQ Sequence 765 BP; 176 A; 217 C; 237 G; 135 T; 0 U; 0 Other;

Query Match 78.3%; Score 18; DB 6; Length 765;

Best Local Similarity 100.0%; Pred. No. 4.7e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GTGGTGGTGGTGACCGTG 20

|||||

Db 575 GTGGTGGTGGTGACCGTG 558

# RESULT 15

ABA05183/c

ID ABA05183 standard; DNA; 765 BP.

XX

AC ABA05183;

XX

DT 04-MAR-2002 (first entry)

XX

DE Murine prion protein PrP coding sequence.

XX

KW Mouse; prion protein; PrP; antiviral; HIV; prion disease; kuru; virucide;

KW antibacterial; neuroprotective; anti-HIV; Creutzfeld-Jakob disease;

KW Gerstmann-Straeussler-Scheinker disease; fatal familial insomnia;

KW bovine spongiform encephalitis; scrapie; ds.

XX

OS Mus sp.

XX

PN WO200183747-A2.

XX

PD 08-NOV-2001.

XX

PF 30-APR-2001; 2001WO-FR001336.

XX

PR 28-APR-2000; 2000FR-00005535.

XX

PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX

PI Leblanc P, Darlix J, Gabus-Darlix C;

XX

DR WPI; 2002-049350/06.

XX

PT New polypeptides, useful as antiviral agents, comprise their prion  
PT proteins able to bind nucleic acid, nucleocapsid proteins, and ligands  
PT for use as antiprion agents.

XX

PS Example; Fig 8; 80pp; French.

XX

CC The present invention relates to normal (PrPc) or abnormal (PrPsc) human  
CC or animal prion proteins which are able to bind to DNA or RNA,  
CC particularly of viral, especially retroviral, origin and to nucleocapsid  
CC proteins (NCP) of human or animal retroviruses. These can be used as  
CC antiviral agents, particularly against human immune deficiency virus  
CC (HIV), and in the treatment of prion diseases including Creutzfeld-Jakob  
CC disease, Gerstmann-Straeussler-Scheinker disease, fatal familial

CC insomnia, kuru, bovine spongiform encephalitis and scrapie. The present  
CC sequence is the murine PrP coding sequence  
XX  
SQ Sequence 765 BP; 176 A; 219 C; 235 G; 135 T; 0 U; 0 Other;

Query Match 78.3%; Score 18; DB 6; Length 765;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTG 20  
| | | | | | | | | | | | | | | |  
Db 575 GTGGTGGTGGTGACCGTG 558

Search completed: March 11, 2004, 04:06:06  
Job time : 268 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2004, 03:24:29 ; Search time 50 Seconds  
(without alignments)  
255.277 Million cell updates/sec

Title: US-10-057-890A-26  
Perfect score: 23  
Sequence: 1 gagtggtggtggtgaccgtgaac 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	% Query Match	Length	DB	ID	Description
	1	18.2	79.1	593	4	US-09-669-751-38	Sequence 38, Appl
	2	18.2	79.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	3	18.2	79.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	4	18	78.3	752	4	US-09-367-572-1	Sequence 1, Appli
c	5	18	78.3	772	4	US-09-367-572-3	Sequence 3, Appli
c	6	18	78.3	900	4	US-09-367-572-2	Sequence 2, Appli
c	7	18	78.3	1322	3	US-09-128-450-27	Sequence 27, Appl
c	8	18	78.3	1322	4	US-09-823-494-27	Sequence 27, Appl
	9	18	78.3	1742	4	US-09-205-258-49	Sequence 49, Appl
	10	17.8	77.4	1599	4	US-09-489-039A-2295	Sequence 2295, Ap
	11	17.8	77.4	1668	4	US-09-543-681A-2994	Sequence 2994, Ap



c	12	17.4	75.7	25603	4	US-09-819-607-3	Sequence 3, Appli
	13	17.2	74.8	1520	4	US-09-620-312D-458	Sequence 458, App
c	14	17.2	74.8	1542	3	US-08-685-558A-8	Sequence 8, Appli
c	15	17.2	74.8	1542	4	US-09-765-449-8	Sequence 8, Appli
c	16	17.2	74.8	1870	3	US-09-071-709-6	Sequence 6, Appli
c	17	17	73.9	66	2	US-08-868-162A-1	Sequence 1, Appli
c	18	17	73.9	3816	4	US-09-976-594-614	Sequence 614, App
	19	16.8	73.0	48	4	US-09-119-507B-89	Sequence 89, Appl
	20	16.8	73.0	48	4	US-08-897-556A-89	Sequence 89, Appl
	21	16.8	73.0	48	4	US-09-547-693-89	Sequence 89, Appl
c	22	16.8	73.0	141	3	US-08-702-870A-12	Sequence 12, Appl
	23	16.8	73.0	306	2	US-08-676-279-39	Sequence 39, Appl
c	24	16.8	73.0	1203	4	US-09-543-681A-1979	Sequence 1979, Ap
	25	16.8	73.0	1877	4	US-09-336-643A-13	Sequence 13, Appl
	26	16.8	73.0	36412	4	US-08-311-731A-132	Sequence 132, App
	27	16.6	72.2	1710	4	US-09-618-425-6	Sequence 6, Appli
c	28	16.6	72.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	29	16.6	72.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	30	16.4	71.3	762	4	US-09-431-887-35	Sequence 35, Appl
c	31	16.4	71.3	1000	3	US-09-128-450-25	Sequence 25, Appl
c	32	16.4	71.3	1000	4	US-09-823-494-25	Sequence 25, Appl
	33	16.4	71.3	1503	4	US-09-328-352-2140	Sequence 2140, Ap
c	34	16.4	71.3	2415	4	US-09-220-132-71	Sequence 71, Appl
c	35	16.4	71.3	2471	4	US-09-919-172-56	Sequence 56, Appl
c	36	16.4	71.3	2471	4	US-09-976-594-71	Sequence 71, Appl
	37	16.4	71.3	4104	1	US-07-998-003A-94	Sequence 94, Appl
	38	16.4	71.3	4104	1	US-08-453-274B-94	Sequence 94, Appl
	39	16.4	71.3	4104	1	US-08-453-695A-94	Sequence 94, Appl
	40	16.4	71.3	4104	1	US-08-268-161A-94	Sequence 94, Appl
	41	16.4	71.3	4104	2	US-08-453-702A-94	Sequence 94, Appl
	42	16.4	71.3	4104	3	US-09-099-639-94	Sequence 94, Appl
	43	16.4	71.3	4104	5	PCT-US93-12588-94	Sequence 94, Appl
	44	16.4	71.3	4104	5	PCT-US95-08071-94	Sequence 94, Appl
	45	16.4	71.3	4650	1	US-07-998-003A-102	Sequence 102, App

#### ALIGNMENTS

#### RESULT 1

US-09-669-751-38

; Sequence 38, Application US/09669751

; Patent No. 6551575

; GENERAL INFORMATION:

; APPLICANT: Greenspan, Ralph J.

; TITLE OF INVENTION: Methods for Identifying Compounds for

; TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to

; TITLE OF INVENTION: Balance and the Perception of Gravity

; FILE REFERENCE: P-NI 3864

; CURRENT APPLICATION NUMBER: US/09/669,751

; CURRENT FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US 60/168,579

; PRIOR FILING DATE: 1999-12-02

; NUMBER OF SEQ ID NOS: 261

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 38

; LENGTH: 593

; TYPE: DNA  
; ORGANISM: Drosophila  
US-09-669-751-38

Query Match 79.1%; Score 18.2; DB 4; Length 593;  
Best Local Similarity 87.0%; Pred. No. 82;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
| | | | | | | | | | | | | | | | | |  
Db 87 GAGTGGTGGTGGTGACCCCGACC 109

RESULT 2

US-09-103-840A-2

; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 79.1%; Score 18.2; DB 3; Length 4403765;  
Best Local Similarity 87.0%; Pred. No. 79;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
| | | | | | | | | | | | | | | | | |  
Db 1284429 GAGTGGTTGTGGTGACCGGGCAC 1284451

RESULT 3

US-09-103-840A-1

; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.

```
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match          79.1%; Score 18.2; DB 3; Length 4411529;
Best Local Similarity 87.0%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy          1 GAGTGGTGGTGGTGACCGTGAAC 23
             ||||| ||||| ||| ||
Db    1284960 GAGTGGTTGTGGTGACCGGCAC 1284982
```

#### RESULT 4

US-09-367-572-1/c

```
; Sequence 1, Application US/09367572
; Patent No. 6593105
; GENERAL INFORMATION:
; APPLICANT: HOLSCHER, Christina
; APPLICANT: BURKLE, Alexander
; TITLE OF INVENTION: PRION PROPAGATION INHIBITION BY DOMINANT-NEGATIVE PRION
; TITLE OF INVENTION: PROTEIN MUTANTS
; FILE REFERENCE: 4121-109
; CURRENT APPLICATION NUMBER: US/09/367,572
; CURRENT FILING DATE: 1999-11-03
; PRIOR APPLICATION NUMBER: PCT/DE98/00429
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 197 05 786.1
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 752
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PrP[delta]H1
US-09-367-572-1
```

```
Query Match          78.3%; Score 18; DB 4; Length 752;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy          3 GTGGTGGTGGTGACCGTG 20
             ||||| ||||| |||||
Db    559 GTGGTGGTGGTGACCGTG 542
```

RESULT 5

US-09-367-572-3/c

```
; Sequence 3, Application US/09367572
; Patent No. 6593105
; GENERAL INFORMATION:
; APPLICANT: HOLSCHER, Christina
; APPLICANT: BURKLE, Alexander
; TITLE OF INVENTION: PRION PROPAGATION INHIBITION BY DOMINANT-NEGATIVE PRION
; TITLE OF INVENTION: PROTEIN MUTANTS
; FILE REFERENCE: 4121-109
; CURRENT APPLICATION NUMBER: US/09/367,572
; CURRENT FILING DATE: 1999-11-03
; PRIOR APPLICATION NUMBER: PCT/DE98/00429
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 197 05 786.1
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 772
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PrP[delta]H1
; OTHER INFORMATION: 92 to 867
US-09-367-572-3
```

```
Query Match          78.3%; Score 18; DB 4; Length 772;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy          3 GTGGTGGTGGTGACCGTG 20
             |||||
Db          563 GTGGTGGTGGTGACCGTG 546
```

RESULT 6

US-09-367-572-2/c

```
; Sequence 2, Application US/09367572
; Patent No. 6593105
; GENERAL INFORMATION:
; APPLICANT: HOLSCHER, Christina
; APPLICANT: BURKLE, Alexander
; TITLE OF INVENTION: PRION PROPAGATION INHIBITION BY DOMINANT-NEGATIVE PRION
; TITLE OF INVENTION: PROTEIN MUTANTS
; FILE REFERENCE: 4121-109
; CURRENT APPLICATION NUMBER: US/09/367,572
; CURRENT FILING DATE: 1999-11-03
; PRIOR APPLICATION NUMBER: PCT/DE98/00429
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 197 05 786.1
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 900
```

; TYPE: DNA  
; ORGANISM: mouse prion  
US-09-367-572-2

Query Match 78.3%; Score 18; DB 4; Length 900;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTG 20  
|||||||  
Db 674 GTGGTGGTGGTGACCGTG 657

RESULT 7

US-09-128-450-27/c  
; Sequence 27, Application US/09128450  
; Patent No. 6211149  
; GENERAL INFORMATION:  
; APPLICANT: Chesebro, Bruce W  
; APPLICANT: Caughey, Byron W  
; APPLICANT: Chabry, Joelle  
; APPLICANT: Priola, Susette  
; TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion  
; TITLE OF INVENTION: Protein  
; FILE REFERENCE: 50121  
; CURRENT APPLICATION NUMBER: US/09/128,450  
; CURRENT FILING DATE: 1998-08-03  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 27  
; LENGTH: 1322  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (101)..(865)  
US-09-128-450-27

Query Match 78.3%; Score 18; DB 3; Length 1322;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTG 20  
|||||||  
Db 675 GTGGTGGTGGTGACCGTG 658

RESULT 8

US-09-823-494-27/c  
; Sequence 27, Application US/09823494  
; Patent No. 6355610  
; GENERAL INFORMATION:  
; APPLICANT: Chesebro, Bruce W  
; APPLICANT: Caughey, Byron W  
; APPLICANT: Chabry, Joelle  
; APPLICANT: Priola, Susette  
; TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion

```
; TITLE OF INVENTION: Protein
; FILE REFERENCE: 50121
; CURRENT APPLICATION NUMBER: US/09/823,494
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/128,450
; PRIOR FILING DATE: 1998-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 1322
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(865)
US-09-823-494-27
```

```
Query Match          78.3%; Score 18; DB 4; Length 1322;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy          3 GTGGTGGTGGTGACCGTG 20
             |||||||||||||||
Db          675 GTGGTGGTGGTGACCGTG 658
```

# RESULT 9

```
US-09-205-258-49
; Sequence 49, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: PZ007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
```

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,915  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,019  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,972  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,916  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,373  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,875  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,374  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,917  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,949  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,974  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,883  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,897  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,898  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,962  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,963  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,877  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,878  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/070,923  
; EARLIER FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15

; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1227  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 1742  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (35)  
; OTHER INFORMATION: n equals a,t,g, or c  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (570)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-205-258-49

Query Match 78.3%; Score 18; DB 4; Length 1742;  
Best Local Similarity 90.0%; Pred. No. 1e+02;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTG 20  
||| |||||:|||||||  
Db 341 GAGAGGTGGYGGTGACCGTG 360

RESULT 10

US-09-489-039A-2295  
; Sequence 2295, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2295  
; LENGTH: 1599  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2295

Query Match 77.4%; Score 17.8; DB 4; Length 1599;  
Best Local Similarity 90.5%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGAAC 23  
||||||| |||||  
Db 487 GTGGTGGTGGTGACAGTTAAC 507



RESULT 11

US-09-543-681A-2994

; Sequence 2994, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 2994

; LENGTH: 1668

; TYPE: DNA

; ORGANISM: Proteus mirabilis

US-09-543-681A-2994

Query Match 77.4%; Score 17.8; DB 4; Length 1668;  
Best Local Similarity 90.5%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGA 21  
|||||  
Db 1221 GAGTGGTGGTGGTGTACGTGA 1241

RESULT 12

US-09-819-607-3/c

; Sequence 3, Application US/09819607

; Patent No. 6686176

; GENERAL INFORMATION:

; APPLICANT: BEASLEY, Ellen et al

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL001078

; CURRENT APPLICATION NUMBER: US/09/819,607

; CURRENT FILING DATE: 2001-03-29

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 25603

; TYPE: DNA

; ORGANISM: Human

US-09-819-607-3

Query Match 75.7%; Score 17.4; DB 4; Length 25603;  
Best Local Similarity 94.7%; Pred. No. 2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GTGGTGGTGGTGACCGTGA 21  
|||||  
Db 3671 GTGGTGGTGGTGACGGTGA 3653

RESULT 13

US-09-620-312D-458

; Sequence 458, Application US/09620312D

; Patent No. 6569662

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyan

; APPLICANT: Chen, Rui-hong

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wehrman, Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Yang, Yonghong

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yunqing

; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei

; APPLICANT: John Tillinghast

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and

; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 784CIP2B

; CURRENT APPLICATION NUMBER: US/09/620,312D

; CURRENT FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1105

; SOFTWARE: pt\_FL\_genes Version 1.0

; SEQ ID NO 458

; LENGTH: 1520

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (271)..(1161)

US-09-620-312D-458

Query Match 74.8%; Score 17.2; DB 4; Length 1520;

Best Local Similarity 86.4%; Pred. No. 2.1e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAAC 23

||||||| ||| | |||||

Db 597 AGTGGTGCTGGAGGCCGTGAAC 618

RESULT 14

US-08-685-558A-8/c

; Sequence 8, Application US/08685558A

; Patent No. 6225081

```

; GENERAL INFORMATION:
;   APPLICANT: SHIMOMURA, Takeshi
;   APPLICANT: KAWAGUCHI, Toshiya
;   APPLICANT: KITAMURA, Naomi
;   APPLICANT: MIYAZAWA, Keiji
;   TITLE OF INVENTION: NOVEL PROTEIN, DNA CODING FOR SAME
;   TITLE OF INVENTION: AND METHOD OF PRODUCING THE PROTEIN
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
;     STREET: 2100 Pennsylvania Avenue, N.W.
;     CITY: Washington
;     STATE: DC
;     COUNTRY: USA
;     ZIP: 20037
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy Disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/685,558A
;     FILING DATE: 24-JUL-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: JPA Hei 7-187135
;     FILING DATE: 24-JUL-1995
;   INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 1542 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: linear
;     MOLECULE TYPE: cDNA to mRNA
;     ANTI-SENSE: no
;     ORIGINAL SOURCE:
;       ORGANISM: Homo sapiens
;       STRAIN: MKN45
;     FEATURE:
;       NAME/KEY: coding sequence
;       LOCATION: 1 to 1542
;       IDENTIFICATION METHOD: by experiment
;       NAME/KEY: signal peptide
;       LOCATION: 1 to 105
;       IDENTIFICATION METHOD: by experiment
;       NAME/KEY: mature peptide
;       LOCATION: 106 to 1542
;       IDENTIFICATION METHOD: by experiment
US-08-685-558A-8

```

```

Query Match          74.8%; Score 17.2; DB 3; Length 1542;
Best Local Similarity 86.4%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 GAGTGGTGGTGGTGACCGTGAA 22
        | ||||| ||||| ||||| |
Db     1457 GGGTGGTGGTGGTGTCCGTGGA 1436

```

RESULT 15

US-09-765-449-8/c

; Sequence 8, Application US/09765449

; Patent No. 6465622

; GENERAL INFORMATION:

; APPLICANT: SHIMOMURA, Takeshi

; KAWAGUCHI, Toshiya

; KITAMURA, Naomi

; MIYAZAWA, Keiji

; TITLE OF INVENTION: NOVEL PROTEIN, DNA CODING FOR SAME  
AND METHOD OF PRODUCING THE PROTEIN

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS

; STREET: 2100 Pennsylvania Avenue, N.W.

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/765,449

; FILING DATE: 22-Jan-2001

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/685,558

; FILING DATE: <Unknown>

; INFORMATION FOR SEQ ID NO: 8

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1542 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ANTI-SENSE: no

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; STRAIN: MKN45

; (ix) FEATURES:

; SEQUENCE DESCRIPTION: SEQ ID NO: 8

US-09-765-449-8

Query Match 74.8%; Score 17.2; DB 4; Length 1542;

Best Local Similarity 86.4%; Pred. No. 2.1e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAA 22

| ||||| ||||| ||||| |

Db 1457 GGGTGGTGGTGGTGTCCGTGGA 1436

Search completed: March 11, 2004, 08:08:34

Job time : 62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2004, 04:03:04 ; Search time 938 Seconds  
(without alignments)  
90.274 Million cell updates/sec

Title: US-10-057-890A-26  
Perfect score: 23  
Sequence: 1 gagtggtggtggtgaccgtgaac 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description

	1	23	100.0	23	14	US-10-057-890A-26	Sequence 26, Appl
c	2	23	100.0	82	14	US-10-057-890A-21	Sequence 21, Appl
	3	22	95.7	79	14	US-10-057-890A-20	Sequence 20, Appl
c	4	19	82.6	1211	12	US-10-425-114-30163	Sequence 30163, A
	5	18.8	81.7	539	12	US-10-424-599-109140	Sequence 109140,
c	6	18.4	80.0	952	12	US-10-424-599-118532	Sequence 118532,
	7	18.4	80.0	2295	9	US-09-799-777-144	Sequence 144, App
c	8	18.4	80.0	114793	14	US-10-148-806-3	Sequence 3, Appli
	9	18.2	79.1	44	15	US-10-455-695-2	Sequence 2, Appli
	10	18.2	79.1	69	15	US-10-455-695-8	Sequence 8, Appli
	11	18.2	79.1	499	15	US-10-027-632-288796	Sequence 288796,
	12	18.2	79.1	593	14	US-10-255-536-38	Sequence 38, Appl
	13	18.2	79.1	1035	15	US-10-260-238-272	Sequence 272, App
	14	18.2	79.1	1860	14	US-10-156-761-429	Sequence 429, App
	15	18.2	79.1	9025608	14	US-10-156-761-1	Sequence 1, Appli
	16	18	78.3	268	12	US-10-424-599-94882	Sequence 94882, A
c	17	18	78.3	419	12	US-10-424-599-106283	Sequence 106283,
	18	18	78.3	425	12	US-10-424-599-89071	Sequence 89071, A
	19	18	78.3	471	15	US-10-260-238-112	Sequence 112, App
c	20	18	78.3	765	13	US-10-106-574-1	Sequence 1, Appli
c	21	18	78.3	765	13	US-10-106-574-2	Sequence 2, Appli
c	22	18	78.3	765	13	US-10-106-574-3	Sequence 3, Appli
c	23	18	78.3	765	13	US-10-106-574-4	Sequence 4, Appli
c	24	18	78.3	1320	13	US-10-115-984-1	Sequence 1, Appli
c	25	18	78.3	1371	12	US-10-425-114-11773	Sequence 11773, A
	26	18	78.3	1659	12	US-10-424-599-89149	Sequence 89149, A
	27	18	78.3	1742	10	US-09-933-767-49	Sequence 49, Appl
	28	18	78.3	1742	14	US-10-023-282-49	Sequence 49, Appl
c	29	17.8	77.4	250	12	US-10-424-599-90636	Sequence 90636, A
c	30	17.8	77.4	350	12	US-10-424-599-75661	Sequence 75661, A
	31	17.8	77.4	445	9	US-09-783-590-11454	Sequence 11454, A
c	32	17.8	77.4	1147	12	US-10-240-425-1446	Sequence 1446, Ap
	33	17.8	77.4	1548	15	US-10-369-493-40512	Sequence 40512, A
c	34	17.8	77.4	1719	15	US-10-369-493-42272	Sequence 42272, A
c	35	17.8	77.4	1764	15	US-10-260-238-700	Sequence 700, App
c	36	17.8	77.4	1821	12	US-10-412-699B-1577	Sequence 1577, Ap
c	37	17.8	77.4	1821	15	US-10-374-780A-1489	Sequence 1489, Ap
c	38	17.4	75.7	253	9	US-09-864-761-20903	Sequence 20903, A
c	39	17.4	75.7	454	9	US-09-864-761-4149	Sequence 4149, Ap
c	40	17.4	75.7	617	12	US-10-424-599-105050	Sequence 105050,
c	41	17.4	75.7	672	14	US-10-156-761-2158	Sequence 2158, Ap
	42	17.4	75.7	689	12	US-10-424-599-77020	Sequence 77020, A
	43	17.4	75.7	1390	14	US-10-017-161-1781	Sequence 1781, Ap
	44	17.4	75.7	1390	15	US-10-292-798-1437	Sequence 1437, Ap
c	45	17.4	75.7	2685	15	US-10-094-749-981	Sequence 981, App

#### ALIGNMENTS

##### RESULT 1

US-10-057-890A-26

; Sequence 26, Application US/10057890A

; Publication No. US20030044901A1

; GENERAL INFORMATION:

; APPLICANT: Coleman, Timothy

```
; APPLICANT: Mansfield, Brian
; TITLE OF INVENTION: Scaffold Fusion Polypeptides, Composition for Making the
Same, and Methods
; TITLE OF INVENTION: of Using the Same.
; FILE REFERENCE: PF537
; CURRENT APPLICATION NUMBER: US/10/057,890A
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/265,782
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,858
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 32
; SEQ ID NO 26
; LENGTH: 23
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic oligonucleotides used to join DNA fragments
US-10-057-890A-26
```

```
Query Match          100.0%; Score 23; DB 14; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 GAGTGGTGGTGGTGACCGTGAAC 23
          |||||
Db      1 GAGTGGTGGTGGTGACCGTGAAC 23
```

# RESULT 2

US-10-057-890A-21/c

```
; Sequence 21, Application US/10057890A
; Publication No. US20030044901A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy
; APPLICANT: Mansfield, Brian
; TITLE OF INVENTION: Scaffold Fusion Polypeptides, Composition for Making the
Same, and Methods
; TITLE OF INVENTION: of Using the Same.
; FILE REFERENCE: PF537
; CURRENT APPLICATION NUMBER: US/10/057,890A
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/265,782
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,858
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 32
; SEQ ID NO 21
; LENGTH: 82
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic oligonucleotides used to join DNA fragments
US-10-057-890A-21
```

```
Query Match          100.0%; Score 23; DB 14; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.7;
```

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
|||||  
Db 23 GAGTGGTGGTGGTGACCGTGAAC 1

RESULT 3

US-10-057-890A-20

; Sequence 20, Application US/10057890A

; Publication No. US20030044901A1

; GENERAL INFORMATION:

; APPLICANT: Coleman, Timothy

; APPLICANT: Mansfield, Brian

; TITLE OF INVENTION: Scaffold Fusion Polypeptides, Composition for Making the Same, and Methods

; TITLE OF INVENTION: of Using the Same.

; FILE REFERENCE: PF537

; CURRENT APPLICATION NUMBER: US/10/057,890A

; CURRENT FILING DATE: 2002-01-29

; PRIOR APPLICATION NUMBER: 60/265,782

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/265,858

; PRIOR FILING DATE: 2001-01-31

; NUMBER OF SEQ ID NOS: 32

; SEQ ID NO 20

; LENGTH: 79

; TYPE: DNA

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Synthetic oligonucleotides used to join DNA fragments  
US-10-057-890A-20

Query Match 95.7%; Score 22; DB 14; Length 79;

Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AGTGGTGGTGGTGACCGTGAAC 23  
|||||  
Db 18 AGTGGTGGTGGTGACCGTGAAC 39

RESULT 4

US-10-425-114-30163/c

; Sequence 30163, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B





```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 118532
; LENGTH: 952
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_78042C.1
US-10-424-599-118532
```

```
Query Match          80.0%; Score 18.4; DB 12; Length 952;
Best Local Similarity 95.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      3 GTGGTGGTGGTGACCGTGAA 22
          |||||
Db      777 GTGGTGGTGGTGACCGAGAA 758
```

# RESULT 7

US-09-799-777-144

; Sequence 144, Application US/09799777

; Patent No. US20020091244A1

; GENERAL INFORMATION:

```
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Corley, Neil C.
; Guegler, Karl J.
; Baugh, Mariah
; Sather, Susan
; Shah, Purvi
```

; TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS

; NUMBER OF SEQUENCES: 154

; CORRESPONDENCE ADDRESS:

```
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
```

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
```

; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/09/799,777
```

```

;          FILING DATE: 06-Mar-2001
;          CLASSIFICATION: <Unknown>
;          PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: US/09/002,485
;          FILING DATE: <Unknown>
;          ATTORNEY/AGENT INFORMATION:
;          NAME: BILLINGS, LUCY J.
;          REGISTRATION NUMBER: 36,749
;          REFERENCE/DOCKET NUMBER: PF-0459 US
;          TELECOMMUNICATION INFORMATION:
;          TELEPHONE: (650) 855-0555
;          TELEFAX: (650) 845-4166
;          INFORMATION FOR SEQ ID NO: 144:
;          SEQUENCE CHARACTERISTICS:
;          LENGTH: 2295 base pairs
;          TYPE: nucleic acid
;          STRANDEDNESS: single
;          TOPOLOGY: linear
;          IMMEDIATE SOURCE:
;          LIBRARY: SCORNOT04
;          CLONE: 2965248
;          SEQUENCE DESCRIPTION: SEQ ID NO: 144 :
US-09-799-777-144

```

```

Query Match          80.0%;  Score 18.4;  DB 9;  Length 2295;
Best Local Similarity 95.0%;  Pred. No. 1.4e+02;
Matches 19;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

```

```

Qy          1 GAGTGGTGGTGGTGACCGTG 20
            ||| |||||
Db          877 GAGAGGTGGTGGTGACCGTG 896

```

RESULT 8

US-10-148-806-3/c

```

; Sequence 3, Application US/10148806
; Publication No. US20030138933A1
; GENERAL INFORMATION:
; APPLICANT: Bai, Chang
; APPLICANT: Metzger, Michael
; APPLICANT: Liu, Xiaomei
; TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NHL, A DNA
; TITLE OF INVENTION: HELICASE
; FILE REFERENCE: 20585P
; CURRENT APPLICATION NUMBER: US/10/148,806
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US00/33065
; PRIOR FILING DATE: 2000-12-09
; PRIOR APPLICATION NUMBER: 60/169,970
; PRIOR FILING DATE: 1999-12-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 114793
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-148-806-3

```

Query Match 80.0%; Score 18.4; DB 14; Length 114793;  
Best Local Similarity 95.0%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTG 20  
||| |||||  
Db 30238 GAGAGGTGGTGGTGACCGTG 30219

RESULT 9

US-10-455-695-2  
; Sequence 2, Application US/10455695  
; Publication No. US20040014034A1  
; GENERAL INFORMATION:  
; APPLICANT: Evans, David H.  
; APPLICANT: Yao, Xiao-Dan  
; TITLE OF INVENTION: Method of Producing a Recombinant Virus  
; FILE REFERENCE: 6580-327  
; CURRENT APPLICATION NUMBER: US/10/455,695  
; CURRENT FILING DATE: 2003-06-06  
; PRIOR APPLICATION NUMBER: US 60/385,886  
; PRIOR FILING DATE: 2002-06-06  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 44  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-455-695-2

Query Match 79.1%; Score 18.2; DB 15; Length 44;  
Best Local Similarity 87.0%; Pred. No. 1.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
||||| ||||| ||| ||  
Db 3 GAGTGGTGGTGGTGATGGTGCAC 25

RESULT 10

US-10-455-695-8  
; Sequence 8, Application US/10455695  
; Publication No. US20040014034A1  
; GENERAL INFORMATION:  
; APPLICANT: Evans, David H.  
; APPLICANT: Yao, Xiao-Dan  
; TITLE OF INVENTION: Method of Producing a Recombinant Virus  
; FILE REFERENCE: 6580-327  
; CURRENT APPLICATION NUMBER: US/10/455,695  
; CURRENT FILING DATE: 2003-06-06  
; PRIOR APPLICATION NUMBER: US 60/385,886  
; PRIOR FILING DATE: 2002-06-06  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 8  
; LENGTH: 69  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-455-695-8

Query Match 79.1%; Score 18.2; DB 15; Length 69;  
Best Local Similarity 87.0%; Pred. No. 1.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
||||| ||| ||  
Db 19 GAGTGGTGGTGGTGATGGTGCAC 41

RESULT 11

US-10-027-632-288796  
; Sequence 288796, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 288796  
; LENGTH: 499  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-288796

Query Match 79.1%; Score 18.2; DB 15; Length 499;  
Best Local Similarity 87.0%; Pred. No. 1.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
||||| |||||||||  
Db 114 GAGTGACCGTGGTGACCGTGAAC 136

RESULT 12

US-10-255-536-38

; Sequence 38, Application US/10255536  
 ; Publication No. US20030087807A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Greenspan, Ralph J.  
 ; TITLE OF INVENTION: Methods for Identifying Compounds for  
 ; TITLE OF INVENTION: Motion Sickness, Vertigo and Other Disorders Related to  
 ; TITLE OF INVENTION: Balance and the Perception of Gravity  
 ; FILE REFERENCE: P-NI 3864  
 ; CURRENT APPLICATION NUMBER: US/10/255,536  
 ; CURRENT FILING DATE: 2002-09-25  
 ; PRIOR APPLICATION NUMBER: US/09/669,751  
 ; PRIOR FILING DATE: 2000-09-26  
 ; PRIOR APPLICATION NUMBER: US 60/168,579  
 ; PRIOR FILING DATE: 1999-12-02  
 ; NUMBER OF SEQ ID NOS: 261  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 38  
 ; LENGTH: 593  
 ; TYPE: DNA  
 ; ORGANISM: Drosophila  
 US-10-255-536-38

Query Match 79.1%; Score 18.2; DB 14; Length 593;  
 Best Local Similarity 87.0%; Pred. No. 1.8e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
 |||||  
 Db 87 GAGTGGTGGTGGTGACCCGACC 109

RESULT 13

US-10-260-238-272

; Sequence 272, Application US/10260238  
 ; Publication No. US20040016025A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Budworth, Paul R.  
 ; APPLICANT: Moughamer, Todd G.  
 ; APPLICANT: Briggs, Steven P.  
 ; APPLICANT: Cooper, Bret  
 ; APPLICANT: Glazebrook, Jane  
 ; APPLICANT: Goff, Stephen A.  
 ; APPLICANT: Katagiri, Fumiyaki  
 ; APPLICANT: Kreps, Joel  
 ; APPLICANT: Provart, Nicholas  
 ; APPLICANT: Ricke, Darrell  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
 ; FILE REFERENCE: 60111-NP  
 ; CURRENT APPLICATION NUMBER: US/10/260,238  
 ; CURRENT FILING DATE: 2002-09-26  
 ; PRIOR APPLICATION NUMBER: US 60/325,448  
 ; PRIOR FILING DATE: 2001-09-26

```

; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 272
;   LENGTH: 1035
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (513)..(513)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (588)..(588)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (591)..(591)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (594)..(594)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (596)..(596)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (608)..(608)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (616)..(616)
;   OTHER INFORMATION: n = any nucleotide
;   FEATURE:
;   NAME/KEY: N_region
;   LOCATION: (699)..(699)
;   OTHER INFORMATION: n = any nucleotide
US-10-260-238-272

```

```

Query Match          79.1%; Score 18.2; DB 15; Length 1035;
Best Local Similarity 87.0%; Pred. No. 1.7e+02;
Matches    20; Conservative    0; Mismatches    3; Indels    0; Gaps    0;

```

```

Qy          1 GAGTGGTGGTGGTGACCGTGAAC 23
             ||| ||||| ||||| |||||
Db          368 GAGCCGTGGTGGTGTCCTGAAC 390

```

```

RESULT 14
US-10-156-761-429
; Sequence 429, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

```

```

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 429
;   LENGTH: 1860
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1860)
US-10-156-761-429

```

```

Query Match          79.1%;   Score 18.2;   DB 14;   Length 1860;
Best Local Similarity 87.0%;   Pred. No. 1.7e+02;
Matches    20;   Conservative    0;   Mismatches    3;   Indels      0;   Gaps      0;

```

```

Qy          1 GAGTGGTGGTGGTGACCGTGAAC 23
             | ||||| ||||| ||||| |||
Db          512 GTGTGGTCGTGGTGACCGTTAAC 534

```

# RESULT 15

US-10-156-761-1

```

; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608

```



; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 79.1%; Score 18.2; DB 14; Length 9025608;  
Best Local Similarity 87.0%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAGTGGTGGTGGTGACCGTGAAC 23  
| | | | | | | | | | | | | | |  
Db 583404 GTGTGGTCGTGGTGACCGTTAAC 583426

Search completed: March 11, 2004, 08:24:17  
Job time : 944 secs